DeepFakes in India

From words for the lies to the faces for the lies

by

Achal Shah

A thesis exhibition presented to OCAD University

in partial fulfillment of the requirements for

the degree of Master of Design in Digital Futures

Exhibition : Experimental Media Studio, Graduate Building, April 12, 2022

Toronto, Ontario, Canada, 2022

Abstract

Social media has shifted the narrative of our public discourse. There is an acceleration in the way we consume and spread information. India, saw a major setback when people chose to believe the news that was shared on Whatsapp, which led to lynchings. Looking at the timeline of deepfakes, their growth, and their application, it is highly speculated that within the near future deepfakes can be used for the same purpose of spreading false information. The thesis explores the connection between fake news and deepfakes in the Indian context, which will illustrate the need for greater digital literacy to prevent the continued acceleration of false information. The practical output is a prototype of a realistic and convincing deepfake that is exhibited to create discourse and advocate for greater digital literacy.

Acknowledgement

I would like to express my most sincere and profound gratitude to everyone who supported me and facilitated the knowledge and resources required to produce this thesis project.

To my thesis supervisors, Adam Tindale and Cindy Poremba for always being available and willing to assist me with their highly valuable insights and knowledge. I am grateful for being able to share this exciting journey with them.

To the digital futures, family who continuously have extended their support. Cheers to us, the batch who made it successfully through the infinite lockdowns and the pandemic.

Table of Contents

List of Figures

List of Tables

1. Introduction

Recent advancements in machine learning technology have allowed for the generation of computer-based autonomous synthetic content such as text, images, audio, and video to seem more realistic and believable than ever before. GANS (Generative Adversarial Network Systems) and Deep Learning, have played a positive role in this synthetic creation. Algorithm-generated media may take various forms, but the one that's getting a lot of attention is deepfakes, which creates synthetic audio and video using GANS and Auto Encoders, unsupervised learning systems which generate an output, algorithmically by calculating the input. There appear to be several possible applications for deepfakes, yet seeing the actual scenarios where it is being applied, it seems that the technology will be used for further nefarious reasons, deploying false information in India. The concern is how this technology may affect people's sense of reality.

This thesis journey was an effort into understanding deepfakes. I assumed the timeline of deepfakes was short, given the first deepfake appeared in 2017 (Ajder et al. 1-27), so it would be easier to scope the area. Little did I know about this rabbit hole. There is a huge amount of work happening around deepfakes, in terms of improving the application, commercializing them, scoping their negative impacts, and implementing newer laws. Artists to technologists, lawmakers, and social media agencies around the globe seem to be involved and are playing their part in discussing the usage and ethics of such powerful algorithms. India, although being a massive global market with a large online audience, does not appear to be engaging in similar discussions. The overarching goal of this thesis is to gain public interest and initiate conversations around deepfakes by speculating what potential impacts they can have on the Indian audience. In India, Facebook was launched in 2006 as an open platform for people as young as 12 and as elderly as 60. Whatsapp was introduced in 2009 with the same accessibility basics. Regardless of age, all of us participated in the race and connected with our peer groups. The platforms enabled simple communication and brought the globe closer together, but as Sherry Turkle points out, we were all Alone Together (Turkle). We all knew one another, yet we were all anonymous to each other, and

soon bad actors began abusing the freedom of the space, which bred antisocial behavior. In 2018, India saw significant negative consequences as a result of fake news but was unprepared.

I see deepfakes as an evolutionary step from fake news. Deepfakes are the faces of those words of false information. My lookalike can be recreated in the form of a video through computer systems, and words can be adjusted to make it seem like I am speaking them. I cannot say the output of the technology is very convincing, but I cannot deny that experts are pushing their way into making an ultra-realistic deepfake. Their workflow doesn't stop at the generation process; rather, different software is involved to generate a realistic deepfaked scenario. There was a time when we understood social media as a place to connect. Later, we moved towards indifferences which led to polarization. For example, Microsoft's robot, Tay, which was an autonomous AI chatbot, was released on Twitter and was given a single task to greet people on the internet and learn from their gestures. Within 8 hours the bot was spilling abuses and propaganda (Vincent). The bot was only learning what we were teaching it. People taught him to hate. Social media is a fairly good place with good people, but the concern is that those good people can become targets of AI synthesis depending on how bad actors use this power.

The thesis is an exploration to find a connection between a particular technology and understand what its impacts can be. The connection between deepfakes is made towards how it can be used to spread false information and what could be its consequences. If the speculation fits within the knowledge, the idea for the next step is to start a positive discourse surrounding this technology. India is not prepared for deepfakes. Yet, I believe in information sharing and with the right intentions and support, we can inform the people about the existence and usage of this technology for the real good but until then the trajectory of deepfakes is quite different. The outcome of the project is based on the similar idea of information sharing, by understanding whether the available open-source technology allows for the creation of a convincing fake identity. The finding helps in developing assurance about the impacts that deepfake could have.

1.1   Research Motivation

The first time I was deceived by a deepfake was in a Whatsapp video post in 2020. The video showed a cabinet minister from India campaigning for a state election later that year (In BJP's Deepfake). The minister ran his campaign in English, a language that is not widely spoken in India. Although the minister is from Bihar, a state in eastern India, he mostly speaks Bihari, it occurred to me that the minister may be studying a new language to broaden his horizons. Later that day, I received a copy of the same film in a different language from another chat group. It was odd to see the same video with all of the parts the same but the language being different. My first inclination was to doubt the video's legitimacy and ascertain whether it was a forgery. Both videos claimed to be synthetic and computer-generated. I have a background in filmmaking. I have worked as a video editor and have dealt with audio and video for films. Everything and I still couldn't figure out why the video I saw in a chat group wasn't real. It was the first of its sort I'd ever seen, something both fake and fascinating. Maybe I wouldn't have realized what deepfakes are if I hadn't seen the second video and questioned the creative output of having the same structure, the fact that the elements of the movies are the same but the language used to communicate is different. When I questioned individuals in 2021 about this technology, I learned that just a few of them had any idea what it was. After showing them examples of how the technology was utilized, some of them reported seeing the video or something similar but in a different context. Somehow, it appeared like the people around me were less engaged in this technology, maybe because I still felt the impacts of deepfakes were speculative. Later in my investigation, I came across another deepfake attack from 2018 where a left-wing female journalist's face was morphed onto a pornographic video (Ayyub). It was unpleasant to see the internet mocking her instead of being supportive.

Videos are an excellent medium for presenting information and distributing knowledge to a larger audience. Because not many people in India are familiar with this technology, it seemed necessary to conduct further research on the subject, and as a member of the technical & futures community, it made more sense to find out if deepfakes have the potential to disrupt the public realm or if they merely exist as another meme generation tool. Investigating the domain of synthetic audio and video generation methods, I discovered numerous algorithms that lead to highly speculative

scenarios of the misuse of publicly open data, and with a little education on how to use these techniques, anyone can deploy them openly to disrupt our online media consumption process, which may possibly accelerate the public's difference in opinions by confirming that what they are claiming is true. I felt the need to do further research on the subject and look at what's going on throughout the world, with this technology.

1.2   Research Questions

1.   How can fake news form a grounding structure to speculate deepfakes as a next threat in the information apocalypse in the Indian scenario?

1.3   Scope

I concentrated my efforts in an area of the fake generation where I could make a significant contribution. I used a pre-existing application that was free to use and would allow me to work alongside the input and output data. One of the reasons I chose the "DeepFaceLab" architecture was that open source tools would allow for easy accessibility, which would potentially be used by everyone. I also limited the creative process by using the face-swapping technique. Given the number of works based on the same generating process, it looked like acquiring information may be fairly straightforward. Furthermore, it seemed as though the deep-faking community was using that method for a reason. I used the word "community" since DeepFaceLab has an active forum where interested folks join to learn about the fake generation process. From the discussions, I discovered that the face-swapping technique is an effective and foolproof approach that can persuade the audience depending on how it is used. We might reject the ability of a text, image, or a single voice to be convincing without solid evidence, as they can easily be altered, but a video with a specific person's face on it would not be regarded as fake and would serve as solid proof. As a result, I feel that deepfakes employing face swaps to spread misleading information are a natural progression toward fake news.

Scanning the field, I discovered that major discussions are taking place all over the world in which media experts, artists, journalists, lawmakers, social media companies, and the AI community are coming together and reframing the context of how this technology can be used for good by addressing legal issues such as identity, misuse, availability, and accountability[1]. It appeared that India was not taking part in such a structure. Since every discussion about deepfakes begins with its application being primarily used in non-consensual pornography, it appeared that the negative implications drew more attention from the media, which strangely had an equal role in framing the literacy around the existence of such technology for the audience. Because false information and the bad impacts of technology circulated quicker than the positive things, I approached the method of distributing knowledge about deepfakes through the lens of fake news and the scenarios it led to.

1.4 Challenge

The main difficulty that needed to be addressed was the creation of facial datasets. I didn't think scraping data from the internet was a suitable option. Facebook sponsored a competition for creating deepfake detectors, and they made a massive dataset of faces public for anybody to use, but I didn't want to work with a dataset of random individuals, and I wasn't sure how, or whether Facebook handled the issue ethically when gathering the data. My investigation into the synthetic media domain eventually led me to MetaHumans, photorealistic digital humans from the Unreal Engine. Because the digital characters are open source, free to use, and computationally produced, I didn't have to deal with any issues regarding identity and theft.

---

[1] syntheticfutures.org

2. Mapping Fakes

This section explains what fake news is and how deepfakes may be used as a method to produce more convincing false information. It lays out the relationship between the Indian audience and their information ecosystem, the underlying false information, and their confirmation, leading to speculation that the effects of deepfakes in India would be the sequel to the effects of fake news. The misleading information was so effectively targeted at the people that it had a role in the (de)formation of the social sphere that was buried beneath the beliefs. The critical examination leads to a knowledge of the continually evolving technology systems and their consequences, which raises the issue of whether people are ready for or even want this.

2.1 Fake News, The Prequel

2.1.1 What is it?

Fake news has come to mean many things. It is referred to as satire, they can be information in the form of propaganda, can be conspiracies, click-bait headlines, deceiving gossip, biased advertisements, or any form of personalized information which is misleading. Fake news is fabricated information that mimics news media content in form but lacks the news media's editorial norms and processes for ensuring the accuracy and credibility of information (Lazer 1094-96). It is primarily produced by individuals who are concerned not with gathering and reporting information to the world, but rather with generating a profit by sharing and spreading messages via social media in the form of false information (Chakrabarti). The term "profit" here does not mean only monetary gain, but rather expands to being benefited by disrupting social discourse or making people believe that the information is true, which it might not be. Generally, false information is crafted to be sensational, emotionally charged, misleading, or made-up which is mostly unverifiable facts and scenarios that hinder our collective abilities to make sense of the world which becomes viral especially when it is targeted at people who are likely to believe it (Al-Rawi 63-79). Since, the internet has enabled a whole new way to publish, share, and consume information and news with very little regulation or editorial standards, information is no longer passively consumed but rather

it is shared, liked, commented on, attacked, and defended in all sorts of different ways by hundreds of millions of people which expands the possibilities of interpretation, sometimes leading to confusion.

2.1.2 Challenge

Platforms and their algorithms have enabled "filter bubbles," in which messages from like-minded actors play repeatedly ("Depth and Breadth"), which sometimes leads to developing trust in all kinds of opinions, which leads to information isolation, and the user tends to believe that false information is confirmed, leading to false beliefs. False beliefs lead to misinformation, and readers unknowingly spread the message without realizing that the facts are incorrect. Reuters and Oxford University's survey found that within the Indian population, Whatsapp, YouTube, and Facebook are the primary sources of news ("Overcoming Indifference"). Despite major issues with disinformation and hate speech and the material shared, which has already resulted in acts of hatred, violence, and public outrage, people tend to believe information shared on those platforms. Another study discovered that a lack of confidence in the news may be influenced by variables other than the news itself, such as a lack of digital literacy (Chakrabarti). The issue with acquiring information online directly from a media platform is that readers share the information they get because they believe it is reliable. The biggest difficulty for internet users is the absence of a clear separation between authentic and fake news. In addition, misinformation differs from regular news in that it reaches individuals through a network that is close to them. The user may receive the same information from multiple people or groups, leading to confirmation of the knowledge. The media is now a web of propaganda, and there are more views than news (Bharali and Anupa 118).

2.1.3 Fake News Scenarios

India, being a massive market for Whatsapp, has more than 390.1 (Dean) million active users, and this number is only increasing making it one of the fastest-growing platforms which is also the main source of driving misinformation (What's Driving). Furthermore, the country has one of the highest rates of forwarded messages, implying that many users are spreading material that they

did not create but came from someplace else (Bengali). The flow of information on social media is constant, which means that someone, somewhere, right now, is ingesting information posted by someone else. On several occasions, social media has been accused of failing to keep an eye on bad actors, and the posts people share, which has resulted in influencing public debate, manipulating people's behavior, and undermining the democratic process. Section 79 of the IT ACT specifies that social media companies are not liable for the posts people share. The issues of legislation, monitoring, and privacy are entangled concepts in and of themselves and are outside the scope of this thesis. When news begins to enrage individuals, misinformed users incite violence not just on social media platforms, but also in real life.

In an instance of vigilante justice in Nagaland, India, on March 5, 2015, a crowd of between 7000-8000 people rushed into a prison, pulled a man detained on suspicion of rape out of the Dimapur Central Jail, stripped him, and lynched him to death in an instance of vigilante justice in Nagaland, India. However, following further inquiry, it was discovered that the individual was not the perpetrator of the crime of which he was accused, as reported by the police afterward. The investigation reports that the man became a victim of consensual sex (Singh). The posts started to appear on social media, falsely accusing the man of being a Bangladeshi who had illegally come to India. The false facts led to people grouping up and planning a coup. One of the posts was titled, "wake up from slumber before they chase us out of our homes". A person also commented that he was not ashamed of what had happened to that man. Different posts shared, accused the man of being an outsider, but the facts prove that he was an Indian citizen. It is not known whether the scenario was a planned act of disinformation, but the fact that it was read, shared, and believed by people was a problem that led to an outbreak. In 2018, another case of lynching occurred in PanjuriKachari in Assam. Two men from Guwahati, the capital of the north-eastern state of Assam, were beaten to death by a crowd using bamboo sticks, machetes, and rocks after being mistaken for child kidnappers. A false story was circulated on Facebook and WhatsApp in India, which led to the deaths (Bali and Desai). The video, which was edited to frame a story and shared, was a part of an information campaign that was released in the neighboring country, Pakistan, to increase awareness about child kidnappings. People from Assam seemed to have believed what they saw on the internet, and rather than verifying the scenario, they decided to take matters into their own

hands, and the consequences have been far from what one could contemplate. Outbursts of violence in Shillong due to communal differences between two groups, the lynching of two men by an angry mob, who presumed the victims to be cattle thieves in Jharkhand, and the death of two people in Assam after being misunderstood as child snatchers by a mob are only some of the instances (Bali and Desai). More than a dozen people became victims of cases of lynching between May and July 2018 alone, where the news were shared mainly through Facebook and WhatsApp (Gowen). This infinite freedom with little accountability led to misuse, mis/dis-information, and therefore fake news. Falsehood diffuses significantly farther, faster, deeper, and more broadly than the truth, in all categories of information (Vosough et al.).

Considering the user base of a platform that allows for people to share their thoughts and ideas, and with its policy of free speech in an end-to-end encrypted space, it makes more sense for governing bodies and media agencies to regulate their policies, and teach the users to distinguish between facts and opinions by improving their digital literacy. Only 38 percent of families in India are digitally literate. 61% in urban regions, compared to merely 25% in rural areas (Mothkoor and Fatima). The Ministry of Electronics and IT defines digital literacy as "the capacity of people and communities to use digital technology for meaningful activities in everyday life circumstances." A person is called digitally literate if they are capable of operating a computer, laptop, tablet, or smartphone, as well as other IT-related products. The policy doesn't do justice in a situation where users do not know how to meaningfully use a technology. Merely understanding technology is not enough in today's age. Rather, knowing the potential the medium holds and the impact it can have on a society is more important. Ethical responsibility is about understanding accountability in today's age of accessibility, which is also an important aspect of literacy. Emphasizing the fact that fake news had convinced the readers into believing false information, we now know how aggressively the Indian audience reacted to it. The reason for reviving the moments through data is to suggest that we need to think about what other technology out there has the potential to disrupt the discourse once again.

3.1 Deepfakes, the sequel

3.1.1 What is it?

Have you seen the 2020 video footage of President Nixon, speaking about the moon-landing and how it was a disaster, 2019 video footage of President Obama swearing during a public service announcement or where Mark Zuckerberg announces that he is deleting Facebook, which attracted 72 million views and led to outrage among viewers who believed the content to be authentic? (Kietzmann et al.). Well, none of those videos are real in the sense, nor did Nixon, Obama or Mark Zuckerberg ever said those things, such are the potential applications of technology. Deepfakes are a form of autonomous audiovisual generation technique that allows people to create realistic simulations of anyone's face, voice, or actions. Individual characteristics can be manipulated using a source's facial data, the data of their voice, or of their body, which can then be used to generate a synthetic version of that being and portray the likeness in a situation they were never really part of. These techniques primarily but not exclusively rely on a form of artificial intelligence known as deep learning and GANs. Deepfakes today are becoming highly commercialized. Communities and forums are a key driving force behind the increasing accessibility of deepfakes along with their open source creation software and applications.

3.1.2 The state of deepfakes

It all started in 2017 when the term "deepfake" was first coined. A Reddit user, u/deepfakes, created a Reddit forum of the same name on November 2nd, 2017. This forum was dedicated to the creation and use of deep learning software for synthetically face-swapping female celebrities into pornographic videos (Ajder et al. 1-27) Later, on February 7th, 2018, the sub was removed concerning how it went against Reddit's policy, but the subreddit already had 90,000 subscribers at that time (Cole). The source codes were already released publicly, which meant that any one of the 90,000 subscribers could have downloaded them. Today, the codes have been compiled into a software package and are available on GITHUB under the name DeepFaceLab, which is one of the most accessible software package to generate synthetic content.

In 2019, the Deeptracelab from Sensity AI, an organization working on developing detection algorithms for deepfakes, conducted research and released an extensive report about the state of deepfakes compared to when it was released in 2017. The report revealed that the total number of deepfake videos online is rapidly increasing, with this measurement representing an almost 100% increase, i.e., 14,678 deepfakes online in September 2019, compared to December 2018, which was about 7,964 (Ajder et al. 1-27). According to their updated 2020 report, the number increased to 52,000 in the first half of 2020 (The Hmm). In their 2019 report, it was found that deepfake pornography accounts for a significant majority of the videos online, which is 96 percent (Ajder et al. 1-27), and the remaining 4 percent is what we see majorly, publicly. Pornographic content exclusively targets women, where the ratio is 100 percent female-based, and for the non-pornographic fakes on YouTube, the majority of the subjects are male, i.e., 61 percent (Ajder et al. 1-27). Deepfakes are almost known for creating content that is believable but unethical, as seen by the stats in the report. To tackle the ethical and implementation issues, different countries, media, organizations, and websites are implementing newer laws.

Users who seem to be interested in creating fakes, on the other hand, appear to be on the rise. The tool allows anyone to create realistic fakes without significant investment, and when we look at the community forums, which are filled with discussions and trials of people creating fakes, we notice an interesting trend in which newer individuals are learning faster from the mistakes of previous creators. Deepfakes generated by the popular face-swapping software, FakeApp in 2018 required a vast quantity of input data. Later the apps available were already less rigorous and more accessible in 2019. Zao, a famous Chinese smartphone application, allows users to freely place their faces in scenes from hundreds of movies and TV shows (Navlakha). Zao just requires a selfie to track facial emotions and head position as raw information. Inserting one's face into a classic movie scene merely takes seconds. As technology for producing high-quality deepfakes becomes more widely available, we should anticipate the creation and distribution of fake material, as well as the accompanying confusion and controversy, to rise. Soon, everyone will be able to select to be the star of their favorite film, maybe alongside their spouses, friends, or coworkers. We may be casted in really unpleasant films, either willingly or unwillingly, or we may be seen and heard speaking things we never spoke. There are a multitude of phone applications accessible on the

Apple App Store that may be able to provide quite convincing results. The premium edition of most face-swapping applications even allows the user to choose and upload an output video of their choice rather than select a video from the application's collection. It would be interesting to see what type of output video the premium members choose.

3.1.3   The Tension

Though phone applications make it simple to create an output, they are restricted to merely face swaps. The open-source software offers a considerably broader variety of possibilities for creating fakes. This is a cause for worry since deepfakes target every component of human image production, meaning an individual may be generated and modified in a variety of ways.

Type	Description
Audio Deepfake	
Voice swap	Changing the voice of the speaker with someone else or recreating their voice
Lip-syncing	Manipulate the lip movement, matching them to the speaking words in the video
Video Deepfake	
Face swap	Replacing/Swapping the face of people in the video
Full Body Puppetry	Transposing the movement from one person's body to that of another

Table 1: Different types of deepfakes

Videos have long served as a means of capturing reality. They are a type of record that may be kept and archived for future generations to look back on and remember what happened in their history. Previously, the concern with video editing software was that individuals could not only capture but also modify reality based on how they viewed it. The forecasts were right (Ethics and Synthetic Media). At the very least, we may find several examples of cheap fakes in the context of video fakes. Cheap fakes are those in which the information is manipulated using video editing techniques to appear to be something different. Techniques such as eliminating frames from a video to make a person sound, walk sloppy, fast-forwarding the frames to make it appear like the individual could be exerting force, merging two distinct movies to make it look like one, and changing the concept are just a few examples. However, video specialists have come to the rescue, as they have been able to analyze the videos, compare them to their original source, and discover the inconsistencies amongst them.

The problem with deepfakes is that there will be no source footage to compare them against. They may be created from multiple different datasets. The video itself would be the only proof of the film's existence. The problem is that we have different styles for a fake generation. All compiled will produce a convincing fake, and when tagged along with other techniques of video composition and scene building, we will be able to see a full detailed story with a fake in it. As the internet implies, they are persuasive. We'd have a whole new archive, alongside the real ones, on the internet. Deep-fake videos may show someone trashing things in a drunken frenzy. They may show someone stealing from a business, hurling horrible religious and racist terms, doing drugs, or engaging in any other antisocial or humiliating conduct, such as speaking nonsensically. The consequences might be disastrous depending on the circumstances, timing, and distribution of the fake. In certain cases, proving the hoax may arrive too late to undo the initial damage. Even if the video is eventually shown to be a fake, the consequences may be irreversible (Rini). Also, it would be difficult for individuals in the future to obtain accurate information. When comparing video technicalities between now and history, things are not as difficult to differentiate, but as AI becomes more powerful and can recolor all the black and white frames, and upscale all those pictures, a new era would be opened for content manipulators. We'd be able to rewrite our history. Things would become more complicated if someone could claim that a real film was faked, a

phenomenon known as the "liar's dividend." When a videotape of a Malaysian cabinet minister was leaked in which he seemed to be participating in an intimate act with another guy, the minister disputed the facts as inaccurate and referred to the film as altered (Toews). This might become a typical problem with deepfakes.

While technology seems to be moving at a fast pace, we do not seem to be bracing up. For now, generating fakes is a lengthy process, and people seem to be exploring the available tools in the form of phone applications or web-based interfaces, which has allowed for accessibility for non-technical audiences to be a part of it. For now, the majority of the apps enable users to create face swaps as well as full-body puppetry. Face-swaps are convincing enough, and they do seem to have the potential to be misused or defame someone. The admission requirement is just to have a phone with sufficient storage capacity to download the app and, occasionally, initiation fees for a premium plan, which typically allows a user to upload their input image and video to swap with. Furthermore, the option to upload simply a photo eliminates the consent barrier, allowing the user to create a synthetic replica of any living or deceased person. It is also not difficult to get access to one image. Given the popularity of these applications and the ease with which they might spread, both users and creators of these apps need to pay more attention to the situation, about how the application can not only generate laughable content but can potentially be used to create a more serious impact. When it comes to technology, there is always the question of whether, since one is capable of building it, it should still be released to the public.

3.1.4 Deepfake Scenarios

Faraz Arif Ansari, an Indian filmmaker, was subjected to online bullying and intimidation as a result of the Reface phone app. Faraz is homosexual and Muslim, so growing up in India must have been difficult considering their status. In the image, Faraz's face was swapped with that of a female model advertising lingerie, and the photo was discovered on a public site, Instagram, with homophobic and vile remarks targeting his identity. Faraz mentioned that the swapped photo of him is from an interview (Joshi). Another example of a deepfake phishing scam occurred when blackmailers followed a victim on Instagram, scraped their data from their profile, then swapped

them to various porn scenes and demanded a large sum of money. If anyone refused, the scammers threatened the user with posting the fake content on the internet as well as to their friends and family. In one instance, a man's facial data was collected after he picked up a video call from a scammer. The data was later swapped, giving the impression that he was participating in an intimate act (Joshi). The question is about the shame that the victim has to go through, and viewers are unlikely to sympathize with the person.

Using the Reface app myself and from experience of training a model on a data set of 20 seconds, I know that the previous examples are not going to be convincing enough to make people believe them. Yet who justifies this online bullying? Rana Ayub was a victim whose video had been faked, swapped onto an adult film, and was publicly published on Twitter, a site open to all with a limited restriction policy that allows for the sharing of adult content only with a label marking it as sensitive material and expects users to make moral decisions on their own. The video was shared by a political party leader and received a lot of attention from the public. People on the internet identified Rana as a female escort and began approaching her with offers. Rana is a journalist, and people in India are often targeted for their ability to speak freely, but such tools compel people from any background to be cautious and forthright about avoiding expressing things that may not be seen as appropriate to the audience. The culprit has not yet been caught. This is not the only occasion in India where a female was targeted, countless more videos on the internet indicate a scenario in which females, both ordinary people, and public figures, are becoming targets. This was the first deep fake video the country saw (Ayyub).

In another instance, an Indian politician, Manoj Tiwari of the BJP, agreed to have his deepfake produced. On February 7, 2020, a day ahead of the Legislative Assembly elections in Delhi, two videos of Bharatiya Janata Party (BJP) President Manoj Tiwari criticizing the incumbent Delhi government of Arvind Kejriwal went viral on Whatsapp (In BJP's Deepfake). The leader's original video was in Hindi, a language commonly spoken in India. The fakes produced were in different languages, Haryanvi and English, where the lips of the target person were altered in the video to make it seem he was speaking in those languages. This scenario marked the debut of deepfakes in election campaigns in India. The video was shared over 5800 WhatsApp groups depending on the

user's location. In the United States, using deepfakes of political figures in California and Texas before elections is restricted by law, but as India does not have any law for deepfakes under its jurisdiction, the government themselves released it a day before the elections. To date, three different advertisements have been released portraying India's famous actors, whereas deepfakes have been used commercially. Different instances, including deepfakes, have already begun to rise, and it is yet to see what major impact is it going to have on the people.

Summary

It's tough to assume that deepfakes will not be deployed and will not have the same impact on people as fake news did. Deepfakes may still have significantly exploited the nonconsensual pornography arena, but many technologies have evolved in this manner. Pornography has been a driving force in technological progress (Coopersmith 96). This technology, however, will not stop there. Today, most synthetic media are imperfect. Some subtle abnormalities or patterns identify the media as false due to data constraints or processing resources. But the issue is, where will all of this lead? In other words, synthetic media will continue to develop over the next few years until it is physically indistinguishable from the real. Users appear to drive it in a more obvious direction, resulting in confusion. Deepfakes have the potential to be used for good, but until that time comes, false news scenarios serve as a road map into a possible speculative future for fake applications. Blockchain and AI detection algorithms appear to be protective measures (Fraga-Lamas and Tiago M), but they are not foolproof yet. Also, given that people are not fact-checking it, but computers are, the question of what the detector is trained on and whether a human can beat a detector once learned about arises. For the time being, increasing society's understanding of synthetic media can provide a partial solution to the dangers of this technology. It will take time, but in the meantime, one must be careful about what one wants and what their technology does.

3. Horizon Scan

Deepfake creation as a concept is categorized into three aspects. The classification is divided considering the unethical implications of the technology. The first step is about accessible data, which is the initial step in the construction of fakes. It raises worries about how anyone's data may be collected and manipulated to add a new perspective to the original theme. The second section discusses algorithmic autonomy, which is the next step after finding a good match of data. People appear to be creating outcomes in a hybrid output of autonomous computational creativity, realistic appearing humans, as autonomous computer systems and the cognitive creative process appear to be merging. The final segment, my body-your choice, highlights the primary problem with deepfakes. The hybrid output of autonomous creativity depicts synthetic humans in circumstances that a real person would not choose, whilst in reality, the human body is losing its autonomy in the presence of its virtual portrayal. The section provides an outline of the malpractices that have occurred due to the misuse of technology.

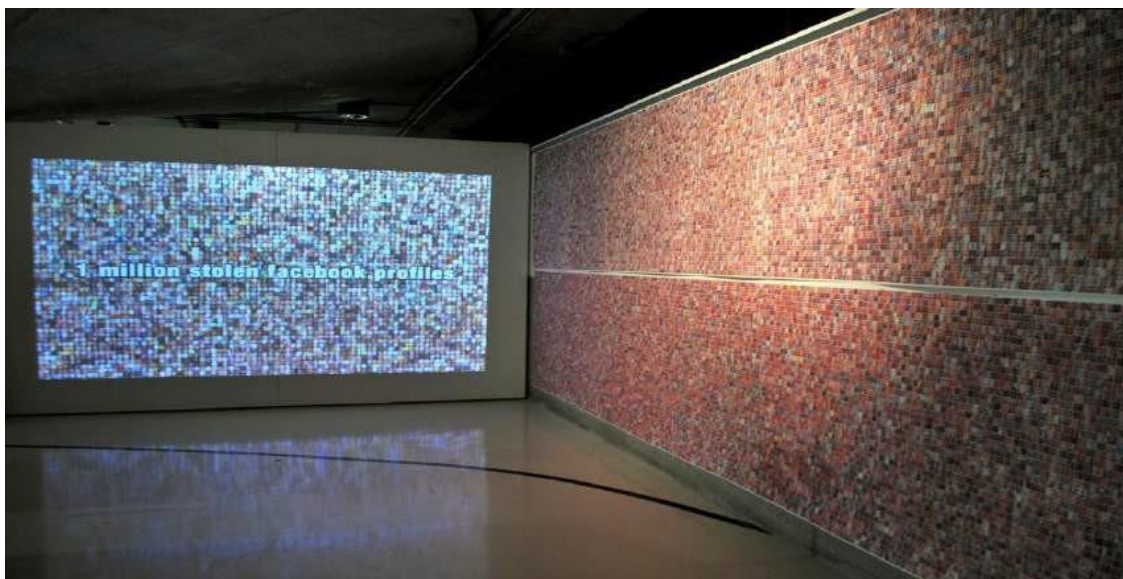## 3.1 Accessible Data

3.1.1 Face to Facebook

Figure 1: Profile photos of stolen facebook profiles

Face to Facebook was created by Paolo Cirio and Alessandro Ludovico. A media artist, Paolo's work focuses on the boundaries of publicly accessible online information, privacy rules, and critiques of commercial and legal information systems. Alessandro Ludovico is an artist, researcher, and chief editor of a neural magazine. The artists created a database of over 1,000,000 Facebook users' public personal data using custom-built software. The database contained information such as name, nationality, profile photo, and relationship status, all of which were obtained without the user's consent. Later, 250,000 profiles were filtered using a custom facial recognition system that used self-learning neural networks to classify users into categories like climbers, easy-going, funny, mild, sly, or smug, with some intuitive differences for male and female participants. Finally, the authors created a custom dating website called lovely-faces.com, importing all 250,000 profiles and carefully personalizing them based on the gathered data.

The purpose was to emphasize the dangers associated with sharing personal information on social networking platforms, and especially the consequences that could occur in real life. The impact of sharing openly and constantly is underestimated since we still have an instinctive tendency to believe that our online activities are constrained within a very personal, visual realm of the screen. Face-to-facebook calls into question internet privacy via one of the web's most recognizable platforms. According to the author, such a move would erode the confidence of Facebook's 500 million members in 2011 (Face to Facebook). If your data is made public, anybody can possess it and reinterpret it in unexpected ways, revealing the internet's vulnerability. Additionally, the stolen and changed data was utilized without the knowledge of Facebook. Interestingly, Facebook is not a dating site, but it does substitute the concepts of a dating website for a few million of its active members, which is why, amongst many things, the authors chose to create a dating website. In the fourth quarter of 2021, there were roughly 2.91 billion Facebook profiles[2], which suggests that the database is much larger than it was before.

---

[2] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Deepfakes need data to swap faces with, and social media provides a huge amount of data for the fake generation. Tomorrow, anyone's accessible data can be used, manipulated, and made to seem like they are hurting the religious sentiments of a group. Since Facebook has no possible way to stop users from downloading each other's data, anyone can become a target now. It is only the creativity of any user that limits them from manipulating the data in any certain way.

3.1.2 Dear Erin Hart

Dear Erin Hart was created by Jessaymyn Lovell and is named after another person named Erin Hart. The artist's wallet was stolen once, and she lost her ID. She then began receiving mysterious bills and summons for court appearances on different small criminal charges and massive amounts of parking penalties. She ultimately discovered that her identity had been stolen and was being used against her. In response, the author hired a private investigator to find the culprit, and once she was identified, the author began trailing her herself, going about her daily life while waiting for her behind tinted windows in an SUV. The author acted as an investigator, victim, stalker, spy, and vigilante. The subject was later confronted, and Erin Heart was sentenced to a year in county jail for charges ranging from Lovell's identity theft case to cheque fraud, forgery, and burglary. Later, the author published the subject's data through an exhibition and also invited her by email, to which she never responded. The artwork emphasizes the fact that human identity, whether physical or digital, is very accessible. What if an individual's identity gets stolen, what if their deepfake is made, not once but again and again and used to propagate false information. What could happen to that individual, and seeing the instances of fake news is not difficult to predict.

3.2 Utilizing Algorithmic Autonomy

3.2.1 Fake Reality

Fake Reality is an interactive audiovisual installation by Paulina Zybinska. The art utilizes the capacity of deepfakes to generate synthetic content. The process begins with collecting data from a willing participant. The individual seems to respond to questions that appear on a screen that

directly shows the personality attributes, which is followed by voice interaction, and the computer then takes a photo of the interactor's face. Based on the interaction, a deepfake video appears. The statements of a participant are edited and synthesized to generate a new version of them. Surprisingly, the modification occurs in real time. The participant will be positioned between the art, which includes several displays and speakers. The result gives me the impression that the information, which is in the form of a movie is being shown on a huge scale, as opposed to what we normally access, is consuming the participant rather than the participant consuming the information. The space allows the viewer to move around the art, which is filled with multiple projections that echo out loud. The work raises questions about the use of autonomous systems, the developer's creativity, and what is achievable with data. It engages the participant in real time, in a physical space, and produces the result in minutes, demonstrating the capabilities of faster systems. Such approaches allow the audience to experience the threat on a very personal level, which, as an art form, is sometimes necessary. As an artist, I am inspired by this approach and intend to apply this method to my work.

3.2.2 This Person Does Not Exist

Figure 2: Faces of synthetic humans generated via GANS

ThisPersonDoesNotExist[3] generates images of human faces that resemble ordinary individuals, but as the domain name implies, those individuals do not exist. The faces on the website are computer-generated, utilizing the GAN (Generative Adversarial Networks) Architecture from the codes released by NVIDIA in a paper called STYLEGAN. Every time the site is reloaded, a realistic-looking but completely fabricated image of a person's face is displayed from scratch. Using a style transfer-inspired alternative generator architecture for generative adversarial networks, an automatically learned model was trained. It was able to generate, unsupervised separation of attributes like poses and identity along with stochastic variation in the generated images like freckles, hair[4]. The new generator enhanced the technical process, resulting in a highly diversified and high-quality collection of human faces. The website was created by Uber software engineer Philip Wang to illustrate the capabilities of artificial intelligence[5]. The output is able to generate realistic human-like looking faces which poses the issue of credibility where in a world where photos and images are the standard confirmation of reality, neural networks automate the process of producing synthetic faces. This image provided by the website is simply downloaded by any user, who can then wear it as a fake identity and create a profile on the internet, which may then be used maliciously. The source code is public, and anybody may train the model to create a specific output in terms of how the fake persons should represent themselves in terms of ethnicity. Who knows if a deepfake with a GAN-generated face gets published in India. What will people focus on: finding the culprit, the information he or she is narrating, or what to do about it?

---

[3] https://this-person-does-not-exist.com/en

[4] https://github.com/NVlabs/stylegan

[5] https://www.inverse.com/article/53280-this-person-does-not-exist-gans-website

3.3 My Body, Your choice

3.3.1 DeepNudes

Figure 3: Screen capture of the DeepNude algorithm's cover page

DeepNudes is a software developed by an unidentified developer. The GAN is activated by giving it a photo of a clothed person, causing it to remove the clothing and render the individual nude. The source code was made available for public usage on the GitHub platform[6]. The tool has only been trained on a dataset of women, so it can only remove the clothing of women, and if the machine is fed a photo of a male, it simply inserts a vulva. DeepNude uses Pix2Pix technology. Using Conditional Adversarial Networks with Image-to-Image translation, an open-source technology was created by researchers at the University of California, Berkeley in 2017 (Corrielus). Image-to-image technology offers a way to fill up image gaps, like how we can remove unwanted things from a photo, and the technology fills in the gaps in the space where objects/subjects are removed. The DeepNudes algorithm essentially washes away the image's unnecessary material with incredibly realistic visuals, like clothes, and modifies them with body parts. DeepNude works

---

[6] https://github.com/yuanxiaosc/DeepNude-an-Image-to-Image-technology

with varying degrees of effectiveness and tends to perform best on images when the subject is already revealing a lot of skin. The 50 dollars paid to make high-definition videos without a watermark is hardly anything if compared to the issue of technology which can generate non-consensual adult content while claiming ownership over women's bodies. The technology can easily be misused to blackmail, defame and silence someone for something they did not do. The Internet seems a little biased towards targeting only females for creating synthetic content within the adult content generation space. However, this raises a concern about portrayal of people in scenarios they never even participated in. We have seen Indian audiences mocking females being represented in such scenarios, which brings utmost shame to the victim as well as their family even when it would be considered a fake.

3.3.2 Epizoo

Marcel Roca is an artist from Spain who uses digital technologies in the fields of mechatronic performance and installation art. His work, Epizoo, is performance art, which enables the spectator to control the performer's body. The whole system comprises a body robot, which is an exoskeleton worn by the performer, a mechanical body-control device, a computer, a vertical projection screen, two vertical lighting rigs, and sound equipment. In the performance the artist stands alone, his nearly naked body is draped in wires fixed to mechanisms that connect to his nose, mouth, buttocks, and pectorals. The mechanism forms as an exoskeletal skin which is activated by the spectator through the touch of a computer. The audience interacts with a character inside a computer in the manner one would in a computer game, which is the control system of the exoskeleton skin. With a click, they can make his mouth twitch, or stretch his lips to impossible lengths. It is the audience's decision whether they want to play around with the performer's body, inflict pain by pushing the limits of his body to near torturous levels, or retreat from the computer. What happens to the artist is projected onto the screen, so that other members of the audience can also see what is happening, leaving the decision to the viewer about whether they want to be a part of it or not. This piece of art is more focused on questioning the audience, even when they have the power to cause harm to someone, considering the tools at their disposal, which are equally powerful enough to do so, what would the audience choose to do. This artwork does not fit into the

realm of digital body manipulation, but it seems to be a very strong piece that raises the same concerns around the issue of body and representation.

Summary

Prior efforts have been restricted to tackling the issues that deepfakes may pose. The objective was to raise awareness about these concerns while also evaluating artworks or applications that focus on their misuse. It's understandable that as autonomous computer systems get more advanced, we want to make our jobs simpler and faster, but we're also gambling on an unknown and occasionally dangerous future. When a web crawler is used, it may steal data at a far faster rate than people. Given the amount of data available on the internet, one might wonder what the next big impact will be. Algorithms are hacked and trained to create harmful information, with premium services starting at $50. Deepfake makers around the internet are also charging the same amount to generate material. It appears that the services are not inexpensive, but this is due to the thinking of those who wish to use them. Fake identities may be manufactured directly by computers and sent into the wild, creating concerns about accountability. On top of that, there are literacy challenges within the audience that play a significant role in dealing with the information they encounter. When the internet first became available, individuals had to educate themselves. Now, similarly, people need to re-learn. In response to deepfakes, Marcel's and Paula's work raised the question of whether artists should make such efforts to make the audience worry about permission, privacy, the body, and the ethics of using sophisticated technology. I similarly plan to focus on the issue of concern and create something similar with my thesis final artwork.

4. Methodologies

4.1 Speculative Design

What exactly is design? Is it a method for finding answers and resolving problems? Is it aesthetic? Speculative Design, as defined by Dune and Raby, is the process of finding problems. Dunne and Raby describe in "Speculative Everything" (Dune and Raby) how a speculative project tries to pose questions about how the world may be, which will make us think about what we want, and then methodologies can be employed to help discover solutions. The goal is to think critically in order to engage in discourse.

This strategy forced me to think critically about the larger implications of deepfakes, and rather than being astounded by the possibilities, the goal was to investigate if and how this technology may impact viewers. General usefulness is still a far-off possibility in the distant future. The hypothesis drove the speculation by comparing it to the results of false news about how anything as basic as texts, video edits, and distorted narratives can confuse people, so why can't a synthetically made film, which would serve as stronger evidence, be used to duplicate the scenarios? A speculation approach is used to generate new ideas and to speculate on how things may be. The goal is to investigate if the hypotheses can act as a catalyst for experience, questions, and sensations, whether happy or unpleasant.

The hypothetical scenario presents a different future, prompting the reader to explore how reality may change. The future may take a different direction than predicted by the thesis, which may be for the better. For the time being, the purpose is to have the reader relate to the problems rather than develop a solution that is either not within the scope or is not possible. The result should encourage the user to pause and evaluate how they could be impacted. It should raise questions like, "Could this happen?" and "Does this reflect the future?" According to Dunne and Raby, speculation as a tool allows art and design projects to conduct research from a critical perspective that isn't tied to a single outcome, but can incite a conversation about the possibilities of an unknowable future in order to create knowledge in multiple and alternative futures (Dunne and

Raby). The research helped form the theoretical idea of making the connection between deepfakes as a method to deploy false information and fake news themselves and speculating why it is important for people to know about this technology.

4.2 Thinking through Making

We can't shape the future if we don't think about it. So, what is the relationship between thinking and making? One makes through thinking and the other thinks through making (Ingold). The process of thinking and making is characterized as a continuous strategy that involves alternating back and forth in order to build newer iterations and by reflecting on what has been done in order to acquire knowledge and insights.

Exploring the road of creation, working with materials and forms, determining what the design is about, anticipating the landscape, improving perception, and creating again are all part of the personal reflection. Thinking is much more than just expressing ourselves verbally, it is also about what we create. Making here is about presenting a narrative via the design of an experience that people can witness for themselves. Hella Jongerius, a designer, compares the process of thinking with her hands to ping-pong, along with the making from the mind, and believes that by combining thinking and making, new sorts of logic and new solutions emerge (Thinking through).

To learn, as Ingold suggests, one needs to watch, listen, and feel by paying attention to the world and what it has to teach (Ingold–Thinking). Gregory Bateson, an anthropologist and cybernetician, calls it deutero-learning (Bateson). This kind of learning aims not so much to provide us with facts about the world as to enable us to be taught by it. The final prototype that I was able to develop, is not an end-step to the search, but raises more questions in the journey of working along with deepfakes. Some of the work does not fall within the scope of this research, but rather it was a continuous process of learning and further thinking about developing iterations and experiences.

5. Prototype

Throughout this thesis project, the idea was to use the available source codes to make face swaps, and the goal was to make them look as realistic as possible. During the making process, different methods were considered, from using apps on the phone to a lengthier process of making fakes with software that allowed for more personalization and manipulation. In search of finding a facial dataset, while considering the boundaries of ethics, I worked with AR, ML, and Blender, but finally settled on utilizing Unreal Engine's MetaHumans. The game character and my own face were then swapped in DeepFaceLab, and the result turned out to be a believable replica of myself. This section talks about the process which helped me understand, what algorithm would work for me and what wouldn't. Each process was about learning while I was exploring the making. It goes into further detail on how I decided to display my creative output to a bigger audience who were unfamiliar with the work I was making during my graduate exhibition at OCAD.

5.1 Experiment One

My first iteration was about learning how to generate fakes. Throughout this procedure, I had a limited grasp of the many types of tools that are accessible. I first saw deepfakes during my first year at OCAD. Because the classes were offered online, I aimed to create a fake that could deliver output in real-time. It all started with the concept of wearing someone else's identity in an online class. I learned how to get around this problem by using Avatirify Python, a standalone piece of software that can be installed on a local computer or run directly on Google Colab, a cloud-based server with its own GPU. The software is built based on the first-order-motion model. The software maps the source's facial data via live video capture and reconstructs the target image in real-time by transferring the video's facial movements on top of the image. There doesn't seem to be a way to merge the vectors that are transferred in the latent space, so instead of a smooth result, there are several stutters and glitches between the video feed and the image. Similarities between the facial features of the output image and the input live video are mapped, which is why there needs to be a facial similarity between the different data.

Figure 4: Left square image is a source input and the right square represents the image which is being manipulated

The image on the left is of me, and it is being streamed live onto the network through a web camera. The picture on the right was produced directly from the website of thispersondoesnotexist. The individual seems to have an open mouth in both frames because the subject's mouth was open when the image was generated. Curiosity drove me to try different picture quality settings to see if the model could track a face on a low-resolution black-and-white photograph or painting. I discovered the model can understand everything until it reads a face, which can then be swapped. To some degree, the output was correct, but the mirroring of facial characteristics such as lips as I was attempting to convey and what it was creating caused the outcome to appear distorted.

5.1.1 Reflection

I concluded that avtarify is not something I would be able to use. Although the concept is still in its early phases, other powerful alternatives are already accessible. The only challenging element of using Avtarify was downloading and running the Python scripts. Aside from that, it was as simple as plugging and playing. One discovery is that deepfakes at this moment must be aesthetically appealing. I think the internet is a really visually appealing medium. At the moment, these twisted, human-like forms do not appear to fit on the internet in such a way that they can persuade. In addition, I needed to create an algorithm that would better map the two datasets.

5.2 Experiment Two

Deep Face Lab is a stand-alone program that may be downloaded and used locally, or it can be configured to run on a cloud-based server. For training reasons, I ran the model on my local PC for training purposes. The program includes more parameters for making fakes, which implies more personalization choices. The algorithm must be fed a diverse set of data, and the AI will go through a massive amount of probability to discover a favorable match. The purpose is to build a facial data library from a source containing various views, expressions, and lighting conditions. The goal is to have a range of face datasets rather than a huge one. The training may take a little longer, but the odds of achieving better results increase. Because of the limitations of my GPU, the native resolution of the facial extraction was set to 256*256. The higher the resolution, the better the switch seems, but 256 was adequate for research purposes. Quick96 was the model that was trained.
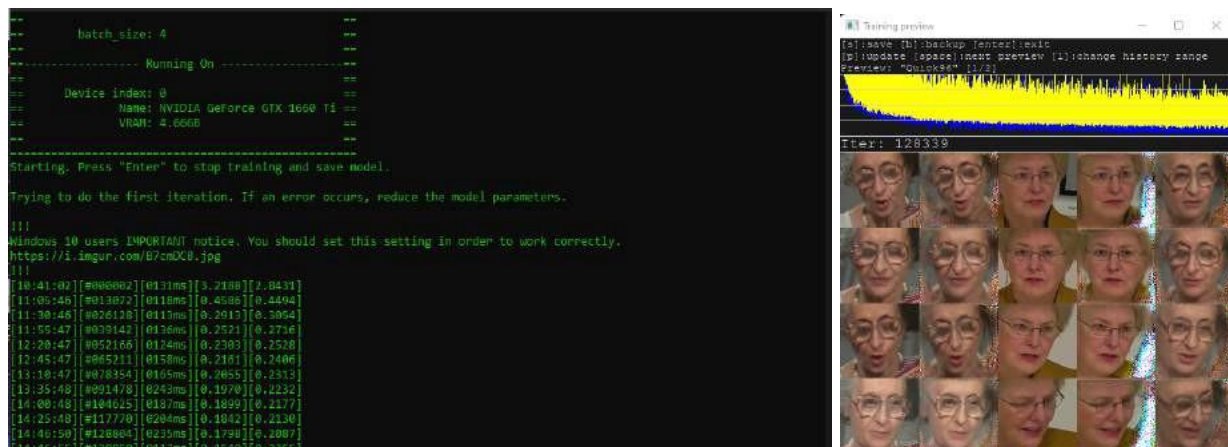
Figure 5: Training of the model in progress, running 120, 000 iterations

Figure 6: Image of the final output after 120,000 iterations

The model was stopped on the third day of training. The results appeared to be satisfactory. Since the video quality was 480p and 1080p, the swapped face appeared a bit blurred. Better resolution properties and longer training will improve the output, but I am concerned about consent issues. I opted to limit my trials to using someone else's face. I realized that creating someone else's fake can also be considered a newer dataset for the next person to train. Later, I worked around existing phone applications. They were simple to use. The technique was simply to upload an input picture and select from the available outputs if using a free version. I had to use another person's face here. The premium version allows users to select the output video from their own library. The Apple play store is flooded with such applications.

Figure 7: Utilizing phone apps to generate fakes

5.2.1 Reflection

It's amazing to think about the possibilities. The only disadvantage has been the simplicity with which face swaps may be made. They're not flawless, and if I swapped my face for Keanu Reeves, people would immediately recognise it as a forgery, but I can still wear his identity and shout abuse online. If I have a premium version, I may choose an output rather than a random film sequence. To generate a swap in the apps, only one picture is required. My question is how tough it is to gain access to anyone's photo, meaning that the user may be ignorant that a fake of them exists. It gets considerably more harmful when paired with voice swapping techniques. As fascinating as these applications are to investigate, it is critical to understand the relevance of online identity portrayal. At the very least, I now understand why I needed to create my own dataset rather than rely on publicly available data.

5.3 Experiment Three

Deepfake's algorithm is based on computer vision, as I discovered. So I began looking at different technologies that could scan facial features and build pictures on top of the existing face. Augmented Reality works on a similar principle of computer vision, although I'm quite sure it's

better at producing animated visuals than realistic human face swaps. I worked with SparkAR and Ebsynth.
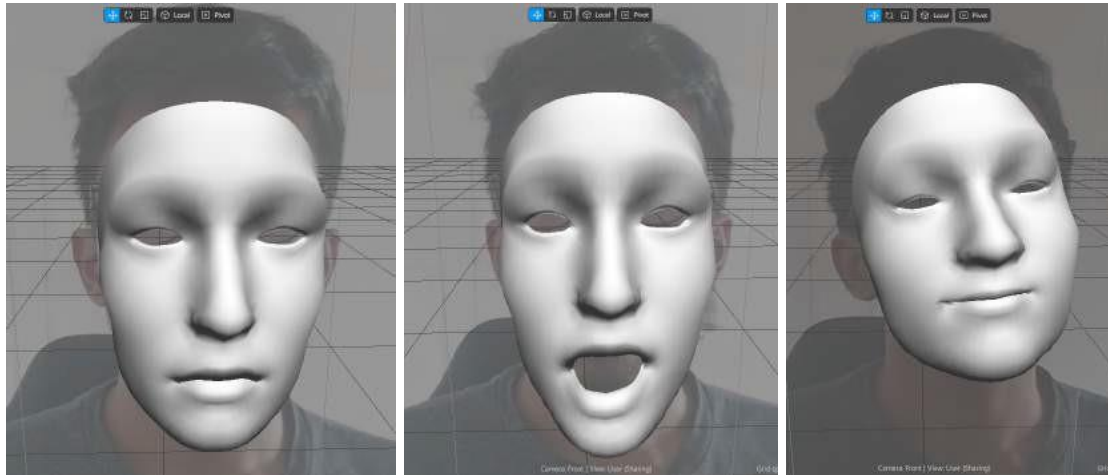

Figure 8: Working with AR based face masks

I applied a face mask in SparkAR. The mask followed the face far better than the Avtarify model. However, it appeared more like a mask and not a skin, which is not what I am attempting to do. I was not able to find the correct blending option. In addition, I had to burn an actor's skin on the mask, which had to be constructed in Photoshop, which extended the procedure even further. It appeared to be a potential method and considering the possibility of uploading AR on Meta's platform, allowing the explorers to test it as an AR piece, it caught my interest even more, but I was eager to work on something less labor-consuming, given the claims that any bedroom producer can develop a deepfake.

At the time, I had learned of another program that was based on the StyleTransfer method. Inside Ebsynth, one could produce the entire video in the same style based on a single input image. It requires two inputs. The first is a video, while the second is a stylized picture. Then it will overlay one onto the other. The underlying architecture is quite similar to that which is applied to construct deepfakes. Because the deepfake algorithm also breaks down a video into several frames, my objective was to construct an animation frame by frame for a data set. Since seeing Scanner Darkly

and Welcome to Chechnya, I've been inspired to employ this strategy. Both films employ a theoretically similar strategy to hide the protagonist's identity.

Figure 9: Creating visualization using Style Transfer

5.3.1   Reflection

With Ebsynth, it is very amusing to see how machine learning is being integrated into the art community. The issue is considerably similar to the previous first-order motion-order model, which was that the software was unable to map the whole space, precisely in every frame of the video, and sometimes the prediction ratio is poor enough that it will generate style swaps in portions not required, which would distort the dataset. Working with AR and Ebsynth, I found their output quite different from the style I wanted to work with. I felt the need to find a more precise technique.

5.4   Experiment Four

At this stage, I did not have any idea about how to generate a customized facial dataset. Continuing the research, I planned on designing visuals surrounding the concept of the body inside the virtual space. I started learning 3D software, believing that it would give me more control over what I wanted to create, and with Blender, being used by major corporations, I believed the learning curve would be convincingly tough but adaptable. Using AI, the first thing I did for the project was to capture an image of myself, and, using that, I created a 3D depiction of myself. I used an AI package

(PuFFHD) to transform myself into a depth picture, then used Blender to convert the obj file into a model. Since I couldn't have access to 3D printers to print a face, I worked within Blender to construct scenarios inside it. The sensation of converting oneself into a three-dimensional figure was strange. I was struck by how empty everything looked to be. It was odd to see myself hollow from the inside when splicing my body in Blender. I liked this workflow, but I wanted to keep trying more things. Considering this method as a potential technique to work on, I continued with my research.
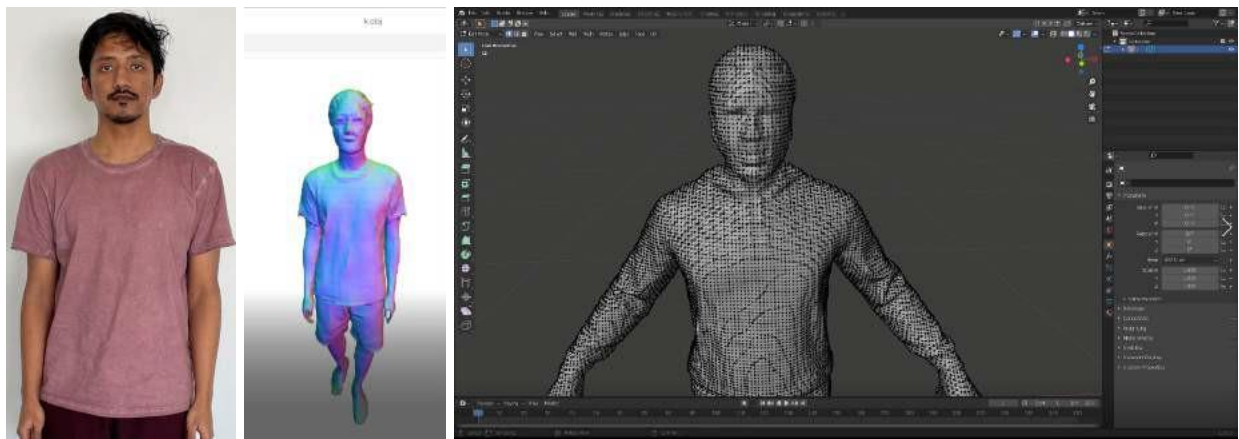
Figure 10: Capturing body and converting the data into a mesh for 3D space
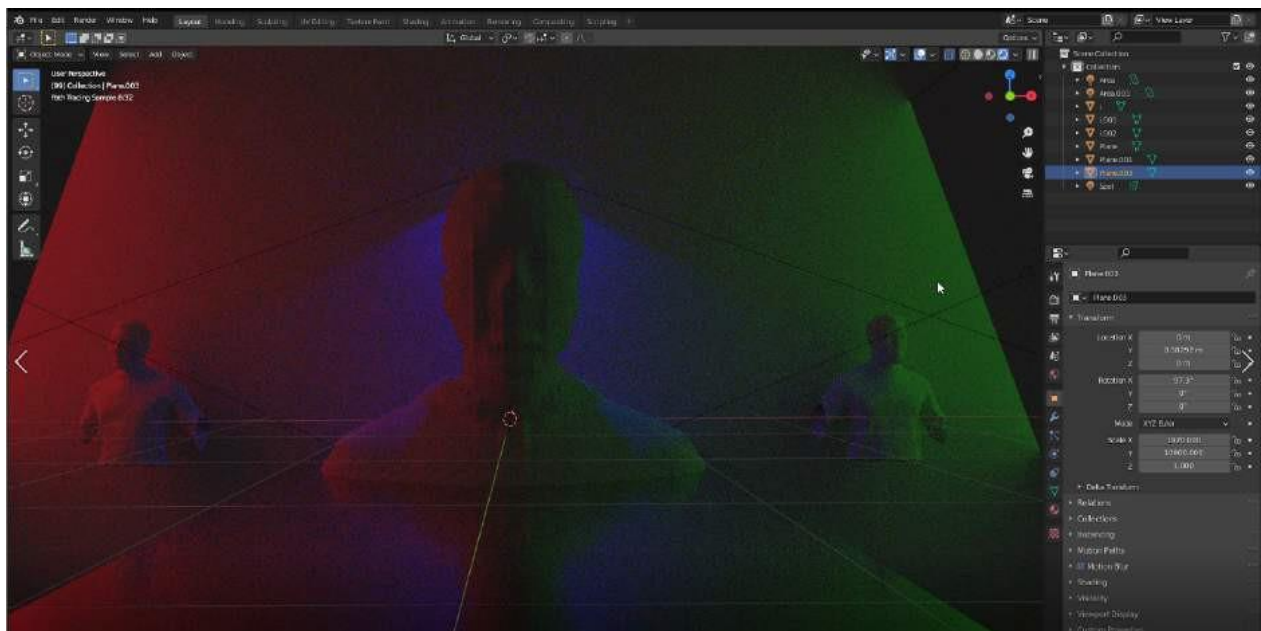
Figure 11: Developing a scene in Blender, using the mesh of my body

5.4.1 Reflection

This stage provided me with an interesting personal insight into the importance of the body. What it means in its representation I enjoyed the outcome, but I saw a lot of promise in other software that looked to be more useful. Blender was a little difficult to understand, but it feels great to go deeper into the 3D space. I was considering creating a short narrative sequence in Blender. However, the investigation of the body in 3D space led me to further my research.

5.5 Experiment Five

While working with 3D engines, I stumbled upon MetaHumans. The Unreal Engine's ultra-realistic gaming characters. The Unreal Engine lets the user create world environments, place people inside those environments, and control the characters' facial expressions. I thought that by utilizing MetaHumans, I might be able to create a facial animation inside a game engine and that dealing with game characters would be easier because all of their movements could be produced inside the computer. After the synthetic individuals are developed on their online platform, the characters will be loaded into the Unreal Engine. The program will then allow me to work on the character's facial expression. The goal was to continue working on the narrative sequence or express the work in a poetic concept and project it into a 3D-printed face of myself using projection mapping after the expressions were mapped. Due to the difficulties in acquiring access to equipment, I was forced to work inside the limits of the virtual world. Through my artwork, I wanted to demonstrate how a person's emotions are captured inside another person's body. As a result, I felt inclined to project MetaHuman onto my face. Furthermore, I did not have access to a projector. As a consequence, I built the scene using Unreal Engine and Blender. The characters and their expressions were created using the Unreal Engine. The face structure and projection visualization were created in Blender.

Figure 12: Generating a MetaHuman

Figure 13: Experimenting with projections inside Blender

5.5.1 Reflection

I admire the approach to this idea, and I was planning on using it as my final output for the presentation. Based on the results of experiments 4 and 5, I intended to continue developing the storyline. It was also a bit simpler to keep control over the landscapes and lights while working on game engines. It occurred to me at that point that the faceset I had made may be considered as another set for face swapping. I decided to swap my face with MetaHumans. All of the processes together lead me to experiment with 3D software and seek a dataset. With each step of the process, I was able to eliminate possibilities that did not seem appropriate for this project. Eventually, it led me to, "Synthetic Humans have no Depth".

6. Synthetic Humans have no Depth

6.1 Overview

"Synthetic Humans Have No Depth" is an ongoing process of transforming a captured reality into a synthetic reality, through algorithmic practices. The process is about creating the characteristics of the body in its absence. The generated intent does not represent the individual's reality but still proves to be a fact about their reality, if believed. In this process of generating alternate realities it felt as though the autonomy and physicality of my body were being lost since the characteristics generated mimicked the real me, but that didn't exactly come from me. Deepfakes appears to be real but aren't. They have the appearance of a human, yet they are not one. They go too far towards a manufactured reality. Seeing myself being generated, I had the impression that the virtual me lacked any sense of taste, smell, touch, hearing, or sight. It doesn't feel like me either. This was most likely the first time I got this close to a character that represented me in 3D space. The entire process of filming myself and superimposing that video onto a computer character, resulting in the final render, watching it, feels like a mirrored reality. How, our other self in the mirror is ourselves, but it is merely a reflection of the actual us until we are present in front of it. The algorithms are, however, so intelligent that once they acquire the knowledge of me in terms of data, they can again and again be coded by anyone, to recreate me in different settings, representing me. Though the recreations would not be the real me and yet all of their acts will have an effect on the actual me. Thus the title.

Figure 14: Generated a deepfake of myself

6.2   Building the Environment

The next stage was to make the MetaHumans look alive after figuring out how to incorporate them into UnrealEngine. When the character is imported, it appears to be a lifelike mannequin. Animating their facial expressions manually, like in experiment 5, is a time-consuming operation. Along with speedier procedures, I required a diverse collection of data to continually try and error those prototypes, feed the model with fresh data, and locate the perfect set to construct a believable fake. A computer vision library (live link) that can be downloaded onto the phone and linked directly to the Unreal Engine is available. The phone and the PC on which the Unreal Engine is operating must have the same IP address, implying a shared network. After connecting the library to MetaHuman, the phone's camera serves as an input, and my facial data is transferred onto the character. The mapping is similar to computer vision, where different data points from my face are linked with the skeleton points of the character. The avatar will mimic my facial motions and does a good job at it. Eventually, I created a dataset based on a gaming character. On the other hand, I went a step farther, filmed myself, and created my facial dataset.
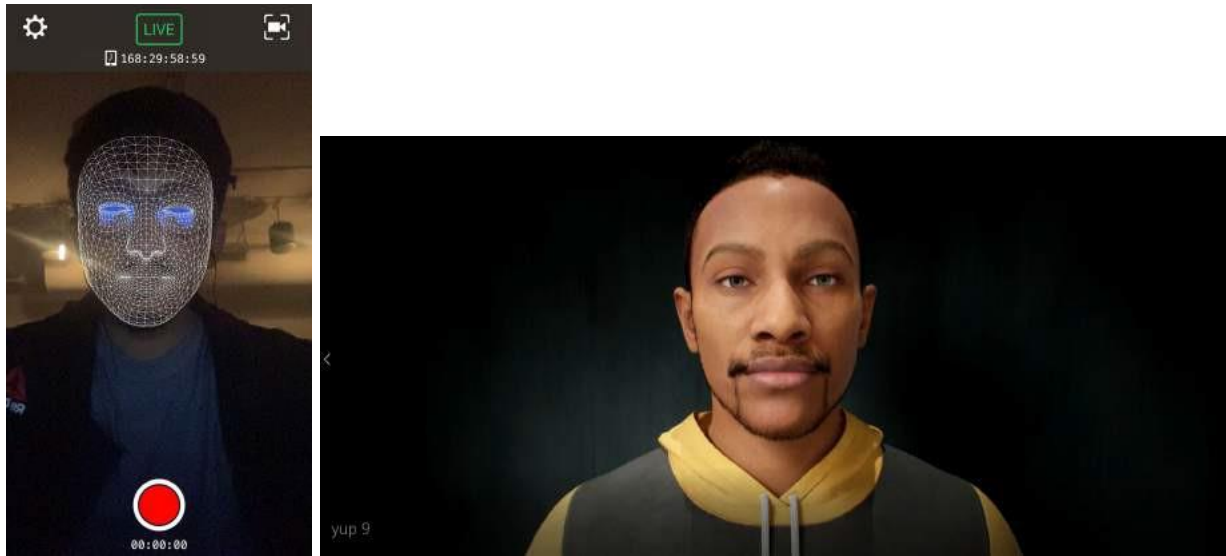
Figure 15: Connecting computer vision library with MetaHumans

The next stage in the approach was to edit the video files in DeepFaceLab. The first stage in creating fakes is to break down each frame of the film into images. As chosen, 24 images were extracted from each frame. The options are 24, 25, and 30 frames. The retrieved images are submitted to the second stage of extraction, in which faces from the photos are extracted. Different resolutions may be used for this extraction, resulting in a high-resolution image of just a face, which also has an impact on the final output. I picked the 128*128 resolution since operating at a higher resolution requires more computing power. Based on our source data, the models may be trained in a variety of contexts. I picked the SAEHD model over Quick96 because it provides a better choice for processing the data. The model provides alpha images of the created fake, which can subsequently be utilized to modify the fake in post-production in After Effects. SAEHD is a large model that will require its section to describe. It is suggested to use this model once the user understands more about training processes. The dataset required cleaning once the face set was extracted from the photographs. Several inconsistencies arise in the obtained data when the AI pulls thousands of photos at once. The dataset can be cleaned manually or automatically by the machine. At this phase, human involvement is essential because the creator only knows what sort of output they want and what dataset will be required to produce it. This method is essential for achieving a high-quality output. If incorrectly extracted data remains in the file, the output may suffer as well. At this point, working alongside the game engine and creating deepfakes seemed

too demanding on the laptop. As a consequence, I started training deepfakes on Google Collab, a cloud platform with GPUs for machine learning. In the end, I trained the model for 300,000 and 600,000 iterations. The final output is trained over a million iterations.

Image 16: Training Process

The model was stopped during its training when the necessary number of iterations was accomplished. The next stage was to combine the faces so that they appeared to be patched together. The merge setting provides various visual enhancing effects and blending choices. After achieving the desired effects, the final step is to export it to a MOV file. Learning to use the program is straightforward, but the challenging case is when various models must be swapped out when training different datasets, and the crucial issue is what sort of parameters are chosen. Although the final product is convincing, there will be some inconsistencies. To detect those flaws, a user may need to learn how to pick out minute characteristic differences around their eyes, ears, or foreheads. They can be extremely visible at times, but most of the time they are quite subtle. Deepfakes. They've reached a convincing level of

realism. They appear to have the capacity to cause confusion between what is real and what is not. The result from the final training shows successful extraction from the dataset. Synthetic Humans have no Depth, aims to provoke feelings of self and how these advancements in technology are going to influence our lives. This metamorphosis is created with the focus of letting people clearly define their virtual presence and know themselves, because from now on, there are going to be multiple fakes that can be generated, creating multiple realities of a single person. Our reliance on technology as a means of writing our path, correcting our mistakes, and guiding us has created many interdependencies, and it has led to humans themselves misusing it. Deepfakes are still in the early stages of development. The output includes certain inconsistencies that may not be visible at first, but what if we let the output and the system exist in its error rather than addressing that mistake? What if, instead of repairing the glitch, we accepted it as a glitch in its own right? This level of realism in a computer is adequate, it does not need to perfectly replicate reality.

6.3   Installation

The final piece was exhibited as an Installation at the Experimental media studio at the OCAD graduate studies building. The exhibition ran for 3 days.

6.3.1   Setting up

The installation required a specific setup, which included projecting the footage onto a customized screen. Cotton cloth-based textiles that were 9ft tall and 1.5 feet wide, black in color were utilized. Six fabrics were suspended from the ceiling. They were placed in a triangle pattern using a progressive grid structure. The objective was to give the picture depth by not putting them on a single piece of fabric. The positioning also allowed people to roam around the space and touch and feel the projections. While creating a deepfake, the program breaks down the pieces of the face into smaller sections and eventually merges all of the different characteristics into a single image. Similar to how the textile arrangement allowed me to split up the pictures, certain portions of my face were only projected onto a portion of the screen. This arrangement was intended to be viewed from various angles, and as the spectator stood entirely in front of the piece, they would see a complete figure of the deepfake. The room was dark, which allowed the projections to be seen. The

projector was placed about 2 feet away from the cloth, which was fastened to a pole. A short-throw projector was employed in particular. The sound was created in Ableton using the glitch and wash libraries. The notion was that deepfakes gave me the impression that there was a malfunction in the system. The sounds' rhythms were a gentle evolution that conveyed the impression of anticipation, and their volume levels rose with time.
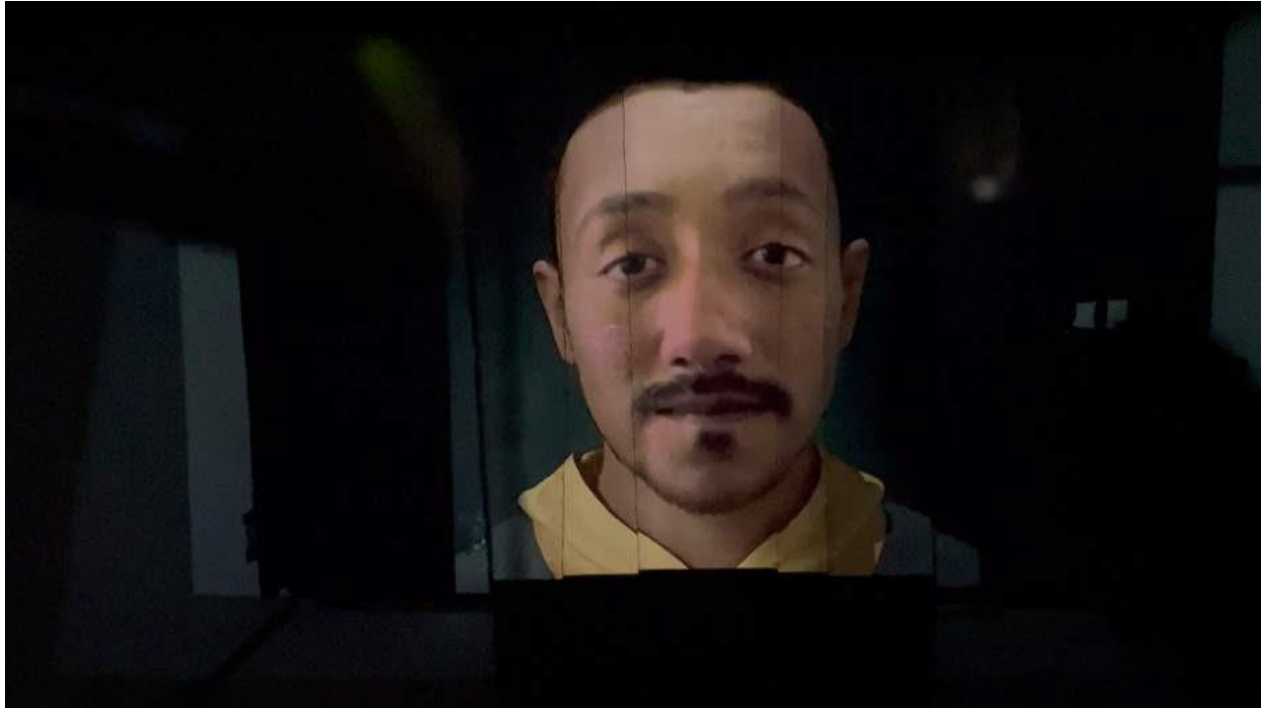

Image 17: Installation from the front

Image 18: Installation from the right

Image 19: Installation from the left-bottom

6.3.2 Response

Several people came to see the show over three days. The overwhelming feedback I received was that people found this deepfake to be quite disturbing. They were certain that the person they were seeing was me. My cohort, with whom I've been working for a longer period, mistook the person on the screen for me. I was hoping for a similar reaction. Many of the people that attended had no idea what deepfakes were. I was delighted to share my knowledge with them. One of the guests mentioned that he wishes he had images of his parents' wedding so that he could relive those experiences. I feel the exhibit was somewhat successful since I was able to persuade individuals, pass on knowledge, and we were able to have a brief discussion.

7. Conclusion

Motivated by the creative vision of investigating if deepfakes have the potential to convey misleading information while remaining realistic enough to trick the viewer, I feel this creative inquiry led me to the conclusion that generating deepfakes is simple. Producing a convincing output is challenging, but not impossible. Based on these facts, I may infer that this technology in the hands of the wrong people will bring a tremendous amount of chaos since the output that can be generated is quite real. The Internet has no limits on who may use the technology, and the commercialization of phone apps considerably reduces the entrance barriers, which are already being used to defame people. Deepface lab and other such software have greater versatility in producing various types of fakes, implying that they might be perceived as more harmful techniques for spreading misleading information. There is widespread worry that the internet will be flooded with a variety of automated materials, resulting in synthetic beings and deceptive information. It is alarming to imagine what deepfakes are capable of accomplishing, as shown in India with how false news can form a social discourse with repercussions that are not restricted to the virtual sphere.

Before making any more breakthroughs, I believe the developer should set specific limits on what kind of content software can be used to generate. The internet is already overflowing with various forms of malicious activity. The manipulation of content is as old as the media itself. People should be adequately educated about the technology's existence and the ultimate output that is triggered by one of these levels of information exchange. While working on the process of creating fakes, I began to perceive the internet as a sphere of free data, and it was quite fun to see how I could construct a large number of fake identities from that perspective. The internet contains a massive collection of people's faces and audio data. I was aware that using those datasets would be deemed very immoral, so I concentrated on creating my dataset. The challenge of how to create my dataset drove me to develop my unique technique for working with game characters. This quest led me to dive deeper into the world of 3D content production, studying Unreal Engine and deepfake software. It was very amusing to learn about 3D space, as well.

Given my position and the expertise I was able to gather over this period of study, about deepfakes, I consider myself fortunate to be able to say that the first time I watched a deepfake, I mistook it for a real video. However, that may no longer be the case today. With the abilities I've acquired, I feel they can be learned by anyone, and with a certain amount of time and dedication, it would be possible for anyone to distinguish between a real and a fake video. Skilled people may therefore be able to help detect and report abnormalities in videos. This would be quite helpful in limiting the flow of false information. However, the question of how much of this effort would be effective in combating disinformation remains unanswered. It is not a detailed plan, but it looks like a solid one at the time. A similar method for teaching people about false news has been investigated in India, where Whatsapp has issued advertisements on the internet and in newspapers. They've been doing a variety of tiny theatrical pieces on the streets for the mobs to gather and learn. Fake news is still a problem, but it is no longer hiding in the shadows. It has been brought to the public's attention. Because knowledge isn't everything, it's also suitable if government regulatory policy becomes legally enforceable. Fighting on two fronts looks to be a feasible strategy until the AI community develops detection algorithms or a block-chain is incorporated into our everyday browsing.

AI-based tools and technologies are expected to advance further, and the algorithms will only provide better representations of humans, but we hope that those who employ these techniques do so for the greater benefit. The more information centered on beneficial use cases, the more people will look at it from a favorable viewpoint. Social media businesses must determine how they will support the publication of deepfakes on their platforms. We also need to understand that people are going to use deepfakes in a variety of ways, ranging from relatively harmless approaches, such as public service chatbots, which generate identities for art or in commercial media, to more destructive scenarios, such as generating and sharing false information about everyone and everything. All of us have to continually evaluate the authenticity of materials, detect harmful applications, and make informed decisions about whether or not to warn the public of the threat. Citizens will need to understand synthetic media as makers, commentators, and sharers, and not only as viewers. The standards they follow in doing so will be critical. According to

Hansen, everything is contingent on a person's ability to embrace, adapt, create, and appraise technology to positively affect their lives, communities, and environment.

7.1 Future Works

The next stage is to find a place, an organization, or an internet support group where I can share this knowledge with a larger audience. Many new organizations are emerging in India that are emphasizing the need for digital literacy. One such organization with whom I wish to be a part is known as the "Internet Freedom Foundation." They are discussing the significance of privacy and data. They have been bringing to light scenes of unethical surveillance methods, data monitoring, and capitalism, and have been battling on the front lines by filing lawsuits. They are conducting interviews with several whistleblowers. In India, nothing like this has ever been done on a wider scale. IFF is disseminating information about the internet infrastructure to the audience. They will have to address the issue of deepfakes sooner or later. When they do, I'd like to share the information I've learnt. IFF has an active forum, and I've seen fewer discussions starting to happen based on deepfakes. I believe they might soon be forming a research team. I am excited to jump on and find out what is happening.

I'm also collaborating with an Indian deepfakes channel, "The Indian Deepfaker '' which has always focused on creating informative satire content. I've always approved of their activities since they've identified their videos as false and are sharing fake footage from across the world to inform Indian people about the scenario and the history of fakes. They've run into some legal concerns as a result of their satirical work and are searching for a solution to build a dataset. They've expressed interest in my approach to using gaming characters, and I'd like to share my expertise with them as well as learn more about their work. For the time being, I'll communicate with them over the internet. We might work on some informative content.

My studies will continue, and I intend to explore techniques to build better fakes and avoid the inconsistencies that I've typically seen. I'm also looking forward to seeing what the film industry does with this technology. Unity will also offer ultra-realistic gaming characters. I'm curious to see

what type of developments they'll make in comparison to Unreal's characters, and if they have a better library, I'll examine if I can produce fakes with it or not. I'm also interested in learning more about body representation in virtual space and might expand my work in that regard. The inventors of the Wav2Lip model, who have achieved a unique breakthrough in lip syncing deepfakes, have given me some encouragement to use their model. Eventually, I'll employ a variety of fake generating techniques, along with face swaps. With everything else, the major focus will be on finding ways to promote awareness and education regarding deepfakes.

Bibliography

Ajder, Henry, et al. *Deeptrace. The State of Deepfakes: Landscape, Threats, and Impact*, 2019, pp. 1–27, www.regmedia.co.uk/2019/10/08/deepfake_report.pdf.

Al-Rawi, Ahmed. "Viral News on Social Media." *Digital Journalism*, vol. 7, no. 1, 2017, pp. 63–79, www.doi.org/10.1080/21670811.2017.1387062.

Ayyub, Rana. "I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me." *HuffPost UK,* 21 Nov. 2018, www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316.

Bali, Aasita, and Prathik Desai. "Fake News and Social Media: Indian Perspective." *Media Watch*, vol. 10, no. 3, 2019, www.doi.org/10.15655/mw/2019/v10i3/49687.

Bateson, Gregory. *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. 1st ed., University of Chicago Press, 2000.

Bengali, Shashank. "How WhatsApp Is Battling Misinformation in India, Where 'Fake News Is Part of Our Culture'." *Los Angeles Times*, 4 Feb. 2019, www.latimes.com/world/la-fg-india-whatsapp-2019-story.html.

Bharali, Bharati, and Anupa Lahkar Goswami. "Fake News: Credibility, Cultivation Syndrome and the New Age Media." *Media Watch*, vol. 9, no. 1, 2018, pp. 118-130, www.doi.org/10.15655/mw/2018/v9i1/49277.

Chakrabarti, Santanu, et al. "Duty, Identity, Credibility: Fake News and the Ordinary Citizen in India." *BBC*, 2018, www.downloads.bbc.co.uk/mediacentre/duty-identity-credibility.pdf.

Cole, Samantha. "Twitter Is the Latest Platform to Ban AI-Generated Porn." *Vice*, 7 Feb. 2018, www.vice.com/en/article/ywqgab/twitter-bans-deepfakes.

Coopersmith, Jonathan. "PORNOGRAPHY, TECHNOLOGY AND PROGRESS." *Icon*, vol. 4, Temporary Publisher, 1998, pp. 94–125, www.jstor.org/stable/23785961.

Corrielus, Fields. "DeepNude: What Image Technology Does DeepNude Involved?." *Gearbest*, 10 July 2019, www.gearbest.com/blog/tech-news/deepnude-what-image-technology-does-deepnude-involved-6727.

Dean, Brian. "WhatsApp 2022 User Statistics: How Many People Use WhatsApp?" *BACKLINKO*, 5 Jan. 2022, www.backlinko.com/whatsapp-users.

Dunne, Anthony, and Fiona Raby. *Speculative Everything: Design, Fiction, and Social Dreaming*. The MIT Press, 2013.

"Ethics and Synthetic Media." Metaphysic, www.metaphysic.ai/ethics-and-synthetic-media.

"Face to Facebook - Hacking Monopolism Trilogy.", www.paolocirio.net/work/face-to-facebook.

Fraga-Lamas, et al. "Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality." *IT Professional*, vol. 22, no. 2, 2020, pp. 53–59. Crossref, www.doi.org/10.1109/mitp.2020.2977589.

Gowen, Annie. "As Mob Lynchings Fueled by WhatsApp Messages Sweep India, Authorities Struggle to Combat Fake News." *Washington Post*, 2 July 2018, www.washingtonpost.com/world/asia_pacific/as-mob-lynchings-fueled-by-whatsapp-sweep-india-authorities-struggle-to-combat-fake-news/2018/07/02/683a1578-7bba-11e8-ac4e-421ef7165923_story.html.

"In BJP's Deepfake Video Shared On WhatsApp, Leader Speaks In 2 Languages." *YouTube*, uploaded by NDTV, 20 Feb. 2020, www.youtube.com/watch?v=XbrpkxIfb0M&t=78s.

"Ingold--Thinking through Making." *YouTube*, uploaded by Pohjoisen kulttuuri-instituutti – Institute for Northern Culture, 31 Oct. 2013, www.youtube.com/watch?v=Ygne72-4zyo.

Ingold, Tim. *Making: Anthropology, Archaeology, Art and Architecture*. 1st ed., Routledge, 2013.

Joshi, Shamani. "The Viral Reface App Could Make Our Deepfake Problem Worse." *Vice*, 21 Aug. 2020, www.vice.com/en/article/wxqkbn/viral-reface-app-going-to-make-deepfake-problem-worse.

- - -. "They Follow You on Instagram, Then Use Your Face To Make Deepfake Porn in This Sex Extortion Scam." *Vice*, 7 Sept. 2021, www.vice.com/en/article/z3x9yj/india-instagram-sextortion-phishing-deepfake-porn-scam.

Kietzmann, Jan, et al. "Deepfakes: Trick or Treat?" *Business Horizons*, vol. 63, no. 2, 2020, pp. 135–46, www.doi.org/10.1016/j.bushor.2019.11.006.

Lazer, David M. J. et al. "The Science of Fake News." *Science*, vol. 359, no. 6380, 2018, pp. 1094–96, www.doi.org/10.1126/science.aao2998.

Mothkoor, Venugopal, and Fatima Mumtaz. "The Digital Dream: Upskilling India for the Future." *Ideas For India*, 23 Mar. 2021 www.ideasforindia.in/topics/governance/the-digital-dream-upskilling-india-for-the-future.html.

Navlakha, Meera. "A Viral Chinese Deepfake App Is Sparking a Debate on Privacy." *Vice*, 3 Sept. 2019, www.vice.com/en/article/qvgkb7/a-viral-chinese-deepfake-app-is-sparking-a-debate-on-privacy.

Singh, Vijaita. "Dimapur Lynching: It Was 'Consensual Sex' Not Rape, Says Nagaland Govt Report." *The Indian Express*, 11 Mar. 2015, www.indianexpress.com/article/india/india-others/dimapur-mob-lynching-victim-was-never-raped-nagaland-govt-tells-mha.

"The Hmm ON Deepfakes." *YouTube*, uploaded by The Hmm, 22 Dec. 2020, www.youtube.com/watch?v=14AuimkpcJA.

"Thinking-through-Making." *Lexicon of Design Research*, www.lexiconofdesignresearch.com/lexicon/texts/thinking_through_making.

Toews, Rob. "Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared." *Forbes*, 10 Dec. 2021, www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=6740c8f37494.

Toff, Benjamin, et. al. "Depth and Breadth: How News Organisations Navigate Trade-Offs around Building Trust in News." *Reuters Institute for the Study of Journalism*, 2 Dec. 2021, www.reutersinstitute.politics.ox.ac.uk/depth-and-breadth-how-news-organisations-navigate-trade-offs-around-building-trust-news.

- - -. "Overcoming Indifference: What Attitudes towards News Tell Us about Building Trust." *Reuters Institute for the Study of Journalism*, 9 Sept. 2021, www.reutersinstitute.politics.ox.ac.uk/overcoming-indifference-what-attitudes-towards-news-tell-us-about-building-trust.

Turkle, Sherry "Connected, but alone? | Sherry Turkle." *YouTube*, uploaded by TED, 3 Apr. 2012, www.youtube.com/watch?v=t7Xr3AsBEK4.

Vincent, James. "Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less than a Day." The Verge, 24 Mar. 2016, www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

Vosoughi, Soroush, et al. "The Spread of True and False News Online." *Science*, vol. 359, no. 6380, 2018, pp. 1146–51, www.doi.org/10.1126/science.aap9559.

"What's Driving India's Fake News Problem?" *The Week*, 12 Nov. 2018, www.theweek.co.uk/97720/what-s-driving-india-s-fake-news-problem.