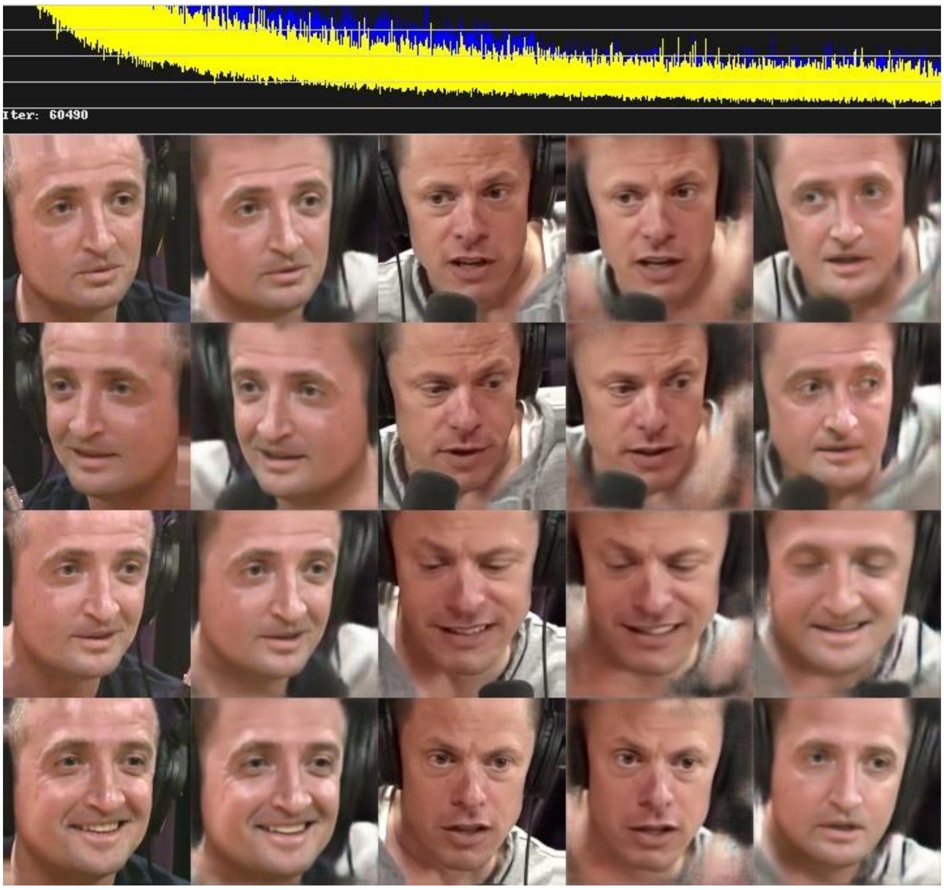
2020

Can Deepfakes be effectively used in VFX?

TECHNICAL REPORT

JOSEPH BISHOP, BA VFX



Abstract

This paper investigates whether it is practical to use Deepfakes in the VFX industry to create digital humans. Deepfakes are a Machine Learning technique which can synthesize an actor's face and overlay it onto a body double. Deepfakes came to prominence in early 2018 when an application called FakeApp was released to the public and has since been used to create various meme videos online.

In this paper I research whether this technique can be used as a cheaper alternative to existing CGI based techniques of creating digital humans. I do this by looking at the theoretical limitations of Auto-Encoder networks and discussing whether Deepfakes can bridge the Uncanny Valley. I will look at existing applications of Deepfakes and discuss how they were achieved. In chapter 2 I look at a tool called DeepFaceLab and investigate whether it can be used in a studio setting. And in chapter 3 I analyse the results from my survey, where I asked participants to distinguish real videos and Deepfakes.

I concluded that currently Deepfakes could have some limited applications in the VFX industry, such as replacing a face in a mid or wide shot, but far more development will be required for Deepfakes to meet or surpass CG.

Table of Contents

Contents

Table of Illustrations

Introduction

In this report, I will be investigating the potential use of Deepfakes in Visual Effects from a technical and ethical perspective. Deepfakes are a Machine Learning technique used to replace one person's face with another person's face. For example, we could replace Harrison Ford's face in *Indiana Jones* with Nicholas Cage's face. This is done by using a Neural Network to learn what the two faces look like and overlaying a synthesized version over the original footage.

Deepfakes were first popularized in early 2018 when a more primitive tool known as FakeApp was released to the public, allowing anyone with the resources to create their own Deepfakes. Following this, Deepfakes were used to create meme videos, and researchers created videos showing how they could be used to create fake videos of politicians. However, there was also a darker side, and a discussion on if we should use people's likeness without their permission.

This technique has the potential to be used in VFX to bring back deceased actors such as Carrie Fisher in *The Last Jedi* or to age or de-age actors. Traditionally this is done by making a 3D recreation of the actor's face and compositing it onto a body double. Due to the level of detail required, this is a very expensive and time-consuming process. Deepfakes can automate part of this process making it faster and more affordable to achieve the same effect. However, the expectations for Deepfakes in VFX that will be viewed on huge cinema screens and 4K TVs are much higher than the Deepfake memes we can view on our phones. The question I will be answering is to what extent can Deepfakes be used in VFX?

In the first chapter, I will look at existing research and the technical side to identify the theoretical limitations of Deepfakes. These include warping artifacts and hardware limitations. I will also discuss whether Deepfakes can bridge the Uncanny Valley. And lastly, I will look at existing applications of Deepfakes in VFX, such as that in *Detective Pikachu*.

During the research process I made several Deepfakes from different angles and focal distances using a tool called DeepFaceLab. In the second chapter, I will write about my own experimentation with a DeepFaceLab and critique the features to see if they could be used in a VFX studio and what needs to be improved.

In the final chapter, I will use my experimentation and feedback to identify the technical limitations of Deepfakes and discuss whether they can be improved in the future. I conducted a survey asking participants to identify Deepfake videos and asked a question about ethics. I've also consulted industry experts for their views. I will also look at whether it is ethical to use Deepfakes to bring back deceased actors from a Machine Learning perspective and compare that to the ethics of using traditional methods to achieve the same result. And using the feedback from my survey and industry research I will conclude my thoughts on the ethics of Deepfakes.

Chapter 1 – Background Research

Fundamentals

Deepfakes work by using a practical application of Artificial Intelligence called Machine Learning. Machine Learning is a wide category of algorithms that try to replicate intelligence by allowing computers to perform tasks through learning instead of being directly instructed to do so. The specific approach to Machine Learning Deepfakes use is called Deep Learning, which tries to achieve this using Artificial Neural Networks. "Deep learning is an approach, it is a structural way of trying to do machine learning, which is a structural way of doing artificial intelligence" (Glassner, 2018). What this means is that Deepfakes are a lot less intimidating than they first seem. They do not use the technology you would see in a science fiction movie yet. They use a very simplified and scaled-down version of how the brain functions.

For VFX this means we cannot just give it the algorithm a video and expect a perfect result. There are limitations we must recognise to learn where Deepfakes can be used, and how things could improve in the future.

Real neurons are too complex to simulate, and we don't fully understand how they work, "…there is chemistry, there is quantum mechanics, there are timing issues, chemicals get released…" (Glassner, 2018) so in Deep Learning, we use extremely simplified versions of neurons. Some people prefer to call them units to reduce media excitement. It takes an input, transforms it in some way and then passes it on to other neurons in the network. The structure of these neurons and the data that they are trained on determine what the network does, this is called the model. (Gerrish, 2018)

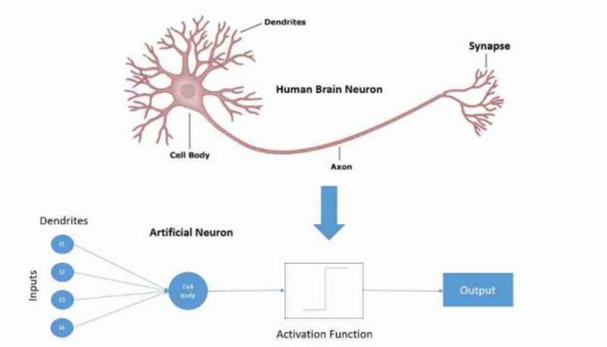


Figure 1: sah, e. (2018). Diagram comparing an organic neuron from an artificial neuron.

Autoencoders

In Machine Learning, a model is structure of the neural network we will use to get our desired result. This model is trained with imagery from the dataset. The training algorithm finds patterns in the training data and adjusts weights in the model, until the desired result is achieved.

To create Deepfakes we need to use Autoencoder networks. An Autoencoder is made up of an Encoder and a Decoder. The encoder compresses the input into a latent-space representation. The decoder aims to reconstruct the input from the latent-space representation. (Zucconi, 2018)

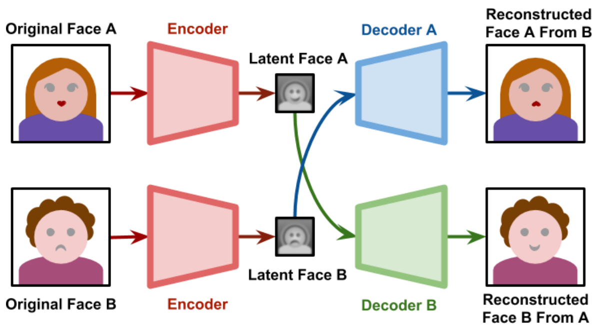


Figure 2: (Zucconi, 2018). Diagram Illustrating a Deepfake network

Each face has an Autoencoder network where they share the same encoder but the decoder is trained differently. After training the network we can put the latent representation from the source and put it into the decoder of the target to swap faces as shown in the graph above.

The problem with using Autoencoders in VFX is that they are inherently lossy. If we train a network where there is the same number of neurons in each layer of the network, the input values are transported to the output nodes without the need to do any learning. "…the network has rewired itself to simply connect the output nodes to the input ones."

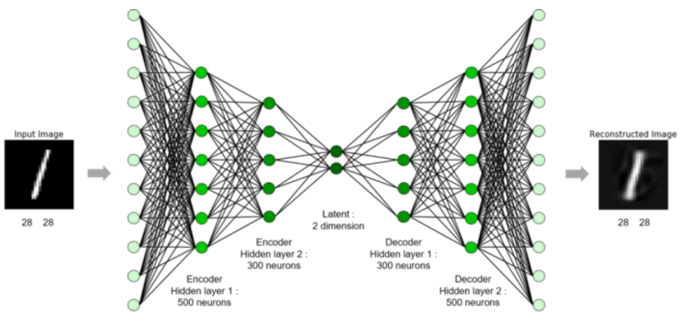


Figure 3: mc.ai (2019). Diagram Illustrating how an Auto-Encoder network compresses images.

(Zucconi, 2018) This means that we need to have a smaller number of neurons in the middle layers, so the network actually learns what the face looks like by compressing it into a smaller space. This is called a bottlebeck and is shown in the middle of Figure 2. While the Decoder does a good job at recovering the compressed data, not everything can be recovered, meaning there will always be some data loss. An example of this is shown in Figure 3 where the input image is sharper than the reconstructed image. For VFX this could be counteracted by overcompensating on the resolution of the face.

This should be possible as many films use workflows where they are shot and worked on in a higher resolution than they are published at. For example the many feature films are shot on the Alexa 65 which has a maximum resolution of 6560x3100 (Fauer, 2014), whereas most films are published at 4K. Technology companies like Apple who will allow filmmakers to work on 8K workflows with the new Mac Pro also enable this. "…Apple Afterburner, a game-changing accelerator card that enables playback of three streams of 8K ProRes RAW video simultaneously. " (Apple, 2019) This means the Deepfakes could be integrated into a scene at 8K and then they will be downscaled to 4K when the movie is published.

Warping Artifacts

Another possible limitation of Deepfakes I've discovered are warping artifacts. When faces are extracted from the source videos, the faces "must undergo an affine warping to match the configuration of the source's face" (Li & Lyu, 2018), this is done so the algorithm only learns about the faces and not the entire image. This means the faces are usually rotated and resampled, depending on its size and position. If a face is too big to fit in the frame it is downscaled, if it is too small it is upscaled.

Studies have shown that this resampling can be detected with certain algorithms (Mire, et al., 2013). Since these algorithms are looking for visual inconsistencies, it is plausible that they could be visible to the viewer. Having more pixels could make this less noticeable to the viewer, so like the previous solution overcompensating on the resolution could solve this issue. Downsampling the overcompensated Deepfake should not be an issue because "Up scaling is generally better detectable than downscaling" (Mire, et al., 2013). This means it more difficult to notice an image has been downscaled than one that has been upscaled.

Detective Pikachu

One application of Deepfakes I have discovered is in *Detective Pikachu*. Deepfakes were used for the first time in a Hollywood movie to de-age Bill Nighy's character Howard Clifford in a short low-resolution news clip. "'We were only doing de-ageing on a few shots so it wasn't worth us building a full computer-generated model of an actor's face,' Mr Webber says." (Bradshaw, 2019). What this shows us is that in some circumstances, such as when a clip is short or low resolution, Deepfakes could be a practical way of achieving similar results as building a 3D model while saving time and money.

Figure 4: (Pokémon Detective Pikachu, 2019) Howard Clifford Deepfake in Detective Pikachu

Since they were using low-resolution video of the same actor, there was very little to go wrong and I can understand why they successfully used this technique. Chris Umé used a similar technique to de-age David Hasselhoff in a news clip (Umé, 2019). He created the Deepfake using old videos of Hasselhoff and using traditional compositing techniques to integrate the face into the scene and add colour to his hair. I think this or a very similar technique was used for the clip in *Detective Pikachu*.

Bridging the Uncanny Valley

One area Deepfakes will excel in is Bridging the Uncanny Valley. "…the cost of a fully digital human is plummeting. The "uncanny valley" is finally being bridged…" (Bradshaw, 2019). In the original essay on the Uncanny Valley, Masahiro Mori argued that the closer a robot comes to looking human the more unsettling it becomes. "…in climbing toward the goal of making robots appear human, our affinity for them increases until we come to a valley…" (Masahiro Mori, 2012). This problem has always plagued Hollywood and is shown in films like *The Polar Express* where some people find the characters look slightly creepy. Studios like Pixar purposely make their characters look unrealistic to avoid this problem (Hsu, 2012), after they discovered that test audiences found *Tin Toy* uncomfortable to watch.

It could be argued that Deepfakes don't have to cross the Uncanny Valley due to their photographic nature, Deepfakes aren't a recreation or replication of the face, they are a representation of compressed facial data which can be applied to a different actor. The network is performing a set of calculations on the photographs it is provided, like using a Premult node just much more complex. As this is a mostly automated process there is no room for human error. This is contrasted to CG, where a complex pipeline of photogrammetry, sculpting, groom and shading etc is required to recreate the same effect from scratch.

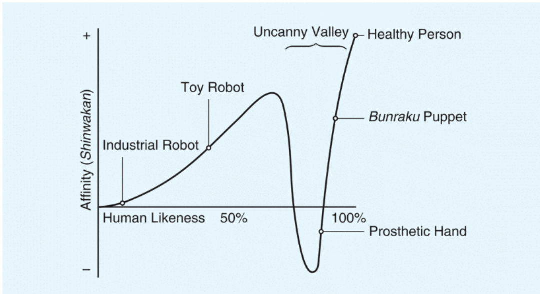


Figure 5: (Masahiro Mori, 2012) Diagram of the Uncanny Valley

Chapter 2 – Creating Deepfakes

DeepFaceLab

To create the test Deepfakes I am using an open source tool called DeepFaceLab (DFL). DFL has several improvements from its predecessor FakeApp, which make it feasible to use in VFX (iperov, 2019). The open source nature of DFL makes it easy to make modifications to the tool, allowing studios to create in-house versions. Instead of a GUI DFL uses batch files which issue Python commands, this makes the software easier to customize while keeping it easy to use. DFL is under active development and improvements and new models are freqeuently added to the software. For example, DFL has a more advanced SAE model which can support resolutions of up to 256px, making it more feasible to create high-resolution close-ups.

For this part of the report, the source I will be primarily focusing on is the DeepFaceLab manual as well as my own experience with the software. This is because it was the most significant source on the subject. The manual was originally written in Russian and it was translated to English using Google Translate, so some of the translations may not be completely accurate.

Models

SAE

The type of autoencoder I will be using is a Sparse autoencoder (SAE). This contains the same features as older models, but it has some new features. For example, we can now transfer the lighting of the face, making it easier to create Deepfakes in different lighting environments.

As of writing this, a new SAE model called SAEHD has been released. Several improvements have been made so the encoder produces a more stable face with less jitter (CyberDainz, 2019). Moreover, CA weights that increase the accuracy of the model are now enabled by default. And a new option called true face has been added. True face should be applied during the last 10k-20k iterations and increases the sharpness in the final result.

Another model that could be used in VFX is the AVATAR model. Instead of swapping somebody's face, the AVATAR model lets you project your facial expressions onto the target. To do this the model firstly learns about the facial differences of the source and the target, secondly the model attempts to recreate the entire scene with the new facial expressions.

Figure 6: (Image by author, 2019) (Livingstone, 2018) Screenshot of SAE training preview

Figure 7: (Image by author, 2019) Screenshot of AVATAR training preview

Training Settings

Resolution

Before you can train the model, you need to extract the faces into a set of square images with the face in the centre. DFL has a facial recognition algorithm that can do this automatically. The algorithm also identifies facial features so the face can be separated from the background. (iperov, 2019)

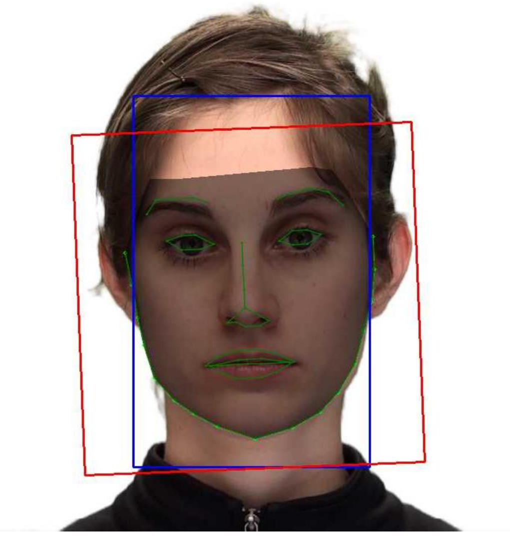


Figure 8: (Image by author, 2019) (Livingstone, 2018) Output of the DFL face extraction algorithm. The facial landmarks show that the algorithm has detected the actors face. The red box shows where the image is cropped to train the model.

The resolution is the size of this extracted image. For VFX the highest possible resolution is desirable, so the result is detailed enough to look Photorealistic.

However, with a high resolution more computational resources are required. A 256x256 image may seem small, but it has 65,536 pixels, which all must be input into the model.

To train a 256px Deepfake it is recommended to have at least 12GB of VRAM. This is an easy requirement for studios to fill, however as most movies are shot in 4K higher resolutions will be required. For example, a 1024x1024 photo would add 1,048,576 values to the neural network.

Using optimizer mode can help, but also utilizing the system RAM. Optimizer mode 2. Allows you to train a model 2x larger, 3. Allows you to train a model 3x larger. However, this will slow down the training process by up to 30%.

Figure 9: (Image by author, 2019) (Livingstone, 2018) Image demonstrating the limitations of using half-face.

You can use the *half face* option to enhance the detail of the face while using the same resolution. This works by cropping in on the face, so the face fills more of the square. This can be useful for low end GPUs or where a very high level of detail is needed, but some parts of the face might be cropped out.

Network Dimensions

In DFL, there are three options to control the number of dimensions in our network; this is the number of neurons in each stage of training.

When the Encoder compresses the face into the latent representation, we can control the number of neurons the Encoder has. Having a high enough number of neurons here is important so the model can record small details in the face, such as small freckles or in the future even skin pores.

The AutoEncoder dims is the number of neurons that store the facial and lighting data in the middle of the network. As mentioned in chapter 1 this layer has less dimensions than the input layer to force the model to learn the facial structure, but within the options, DFL provides a higher number so the model can remember as much detail as possible.

If you have a complex scene, where the actor is making a wide variety of facial expressions and the lighting changes frequently it is important to have a high number here, so the model has enough memory remember everything.

Decoder dims is the number of dimensions there are in the decoder.

In a network with a high resolution but small number of dimensions, problems and artefacts may be noticeable in the model. For example, the face may seem too smooth and lacking in detail compared to other parts of the body. In addition, eyes may not blink properly.

Increasing these settings will have a large effect on the amount of VRAM used, so hobbyists must sacrifice resolution to create photorealistic Deepfakes. This is not desirable in VFX as we are working with 4k+ footage. However, large studios should have the resources to train on professional GPUs with greater amounts of VRAM. I had access to a single Tesla K80 with 12GB of VRAM and 12GB of RAM. A large studio should have the resources to purchase an Nvidia DGX system with multiple Tesla V100's which have 32GB each.

Conversion Settings

Conversion is the final step in the process in which we take our trained model and apply it to the source video. Recently DFL included a new interactive GUI converter which will allow artists to change settings and see the results instantly, rather than relying on trial and error in the command line. This relies on keyboard shortcuts, and there are still things I'd like to be improved, but in theory a VFX studio could make in-house tools to make converting even easier for artists.

Blending

There are a variety of blending modes which can be used to blend the face with the source footage. Firstly, there is the overlay mode, which as the name suggests simply takes the face and overlays it above the source footage. If nothing went wrong during the training process this is usually fine because for an amateur very little can be done to improve the image once it is rendered. But in VFX there may be a lot more to desire. For example, we may want to add grain individually to the face or adjust the mask.

Like overlay there is seamless, which uses a technique called Poisson image editing to blend the together. This seems to work quite well, so I would primarily compare overlay and seamless and see which works best for the image I am working on.

The hist-match option matches the histogram of the source and face footage. On the clips I've used I found this doesn't work so well because raises the black level too much and leaves a white border around the face. Hist-match-bw is a similar operation but only matches histograms in a greyscale channel. I have found this to be better, but in most cases the black level is still too high; this could be corrected in Nuke though.

Finally, there is raw. This option allows you to export the raw unedited Deepfake, which you can export with an alpha channel to edit in compositing software such as Nuke. This is most ideal for VFX studios as it allows artists to use tools that are flexible and that they are familiar with. For example, I could use the Roto tool or blur tools in Nuke to correct the mask and conceal imperfections. While training already does a large part of this, the artist could also apply further colour correction using Nuke's colour nodes.

Figure 10: (Image by author, 2019) Comparison between different blending modes.

Masking

Due to the way Deepfakes work the background is also learnt, so we must use masks to separate the generated face from the background. There are several settings we can use to decide how the matte is generated.

The learned mask is generated from the training process. This will soften some of the edges where necessary but can be quite inconsistent. The FAN masks use a pre-trained model to generate a mask with either the predicted or source face. These results in a smoother mask and can remove items blocking the face from the mask. The learned and the FAN masks can be multiplied together to get the best of both, and I think this works best for me.

Figure 11: (Image by author, 2019) Comparison between different masking modes.

In a studio environment I think these algorithms work well enough to save time, but a Roto artist could make corrections to the matte in Nuke if necessary. Due to the way the background is also learnt the matte doesn't have to be perfect, problems are only very noticeable when the mask is too small, which can be corrected with the erode tool.

Chapter 3 – Findings

During the research process, I have created a set of Deepfakes to help me learn about the strengths and weaknesses of the technique, some of these Deepfakes are included in my survey. When I first started, I challenged myself to create the perfect Deepfake for Hollywood, but I quickly realised this would not be possible with my hardware limitations, so instead I have been trying to make photorealistic Deepfakes at a lower resolution.

In the survey, I created three videos with a mixture of Deepfakes and real clips and asked the participants to identify which clips were Deepfakes. If the participant identified a Deepfake I, then asked they why they noticed it. I did this so I could see how the public would react to Deepfakes and get an unbiased external view on what does and does not work. As shown in Appendix B most of my responses came from a Reddit page called r/SampleSize, which is a community where researchers can share surveys. Most respondents (94%) did not come from a VFX background. I received a total of 103 responses.

Video 1

For my first Deepfake (shown in Appendix A) I tried to swap a face from the RAVDESS database (Livingstone, 2018) with a video I found on YouTube, in order to create a digital double of the YouTuber. The video on the RAVDESS database was a close-up, as I wanted to see how detailed I could make it. I used the maximum resolution of 256px, but I had to reduce the number of autoencoder dimensions to 300. This did not seem to negatively affect the video; I think this is because the lighting stayed the same throughout the shots, meaning less information had to be remembered.

The biggest issue with this video is the blurriness. The problem here is that the resolution of the face is 256x256, but on the video, it is roughly 350x350, meaning some upscaling was done. In addition, the video is 720p, which is not suitable for VFX. At 4K, the size of the face would need to be at least 1000x1000px. This model was only trained to 36,000k iterations. I could have received better results by training up to 80k - 100k iterations.

There is also a visible jitter in this video. I could not find out why this was happening but it could be because I did not use the new SAEHD model which reduces jitter (CyberDainz, 2019). In my newer Deepfakes I have used the SAEHD model, and jitter is much less prevalent.

Because Deepfakes only replicate and overlay the face and no other part of the body, it is important that the double looks familiar to the source actor. I tried to use actors with a similar appearance but I still experienced issues making a convincing digital double. The first issue is with hair, both actors have short hair, but they have slightly different hairlines. The make-up department, who could attempt to replicate the source actor's hairstyle onto the double, could resolve this. The second problem is that the target actor has a wider forehead. This could be solved by using a sibling of the actor, but more research in skull structure and head shape should be conducted to help us find doubles that are more suitable.

Video 2

In Video 2, I used video clips I found from interviews on YouTube. Because this video was a mid-shot rather than a close-up, the Deepfake was sufficiently sharp to be integrated into the target video. However, this Deepfake had a few issues quality, as shown in Appendix B 80.6% of respondents correctly identified the fake. Many respondents answered that this was due to warping, unnatural movement or mismatched colour. When looking at the fake there is an inconsistency in the lighting of the face, one respondent said, "Shadows are weird at times…", and the Deepfake appears to be more diffuse than the forehead. Additionally, there is a problem where the eyes move in an anomalous way.

From my insight and comparisons between other Deepfakes I've made, this happened due to a lack of training data of the source actor. I had a hard time finding videos that shared the same angle, and while I found some, it wasn't enough. Also, much of the lighting was different, most of the source footage appears to be lit directly with a diffuser, whereas the target footage was shot in a studio with overhead lighting.

Overall, this example highlights the importance of having enough training data. For VFX this means either finding a large catalogue of historic footage of the actor, covering a wide range of angles, lighting and facial expressions. Or if possible, using a light stage to capture every angle of the face in different lighting conditions.

Video 3

This video was a very wide shot meaning the face only took up a very small portion of the screen. Due to this it was one of the most successful videos with only 12.6% of respondents selecting the correct answer. Out of the few people who answered this question correctly, 33.3% said it lacked detail, and 22.2% mentioned unnatural movement. When comparing the original face and Deepfake I noticed there is a disparity caused by the compression in the original video. This would not be an issue in industry and is a consequence of using online video, but in later versions I attempted to correct this in Nuke, by denoising the original footage and then renoising it with the face.

This example shows us that Deepfakes could be used in the VFX industry for wider shots where the current limitations of Deepfakes are less noticeable. For example, Deepfakes could be used to make stunt doubles look closer to the real actor.

Video 4

This video was created after conducting my survey, but I found it important to include since it is an improvement on Video 2. To make this video I used two clips from *The Joe Rogan Experience* podcast. I decided this would be a good idea because I was able to create a Deepfake from a controlled environment where the lighting and the camera angles should be the same. In order to preserve facial details, I decided to compromise only resolution by restricting it to 144px so I leave the network dimensions at their default values.

While less detailed this Deepfake retains more facial data compared to Video 2 and has no training problems with eyes or other anomalies. The most noticeable problem with this video is the difference of lighting around the brow ridge, cheeks and nose. Additionally, the Deepfake is too specular in comparison to the forehead. These issues are caused by differences in lighting between both clips.

Whilst I've made improvements over Video 2, this example reinforces the importance of having correct lighting in both clips. While this problem is easy to resolve, it highlights a disadvantage that Deepfakes have in comparison to CG. With CG a face can be captured in neutral lighting in a light stage and lit to match any lighting condition using PBR shaders.

Finally, I asked the respondents if they would have noticed the Deepfakes if they weren't told to actively look for them. 30% of respondents said they would not have noticed, and the rest of the responses were qualitative, with most people saying they would have only seen the first Deepfake as described by this example. "in the first video the second clip was pretty obvious but the other two videos were a lot harder". I think this shows us that while we should be striving for perfection, if made correctly, there are some instances where Deepfakes could be used without being noticed by the average viewer.

Ethics

The debate around recreating somebody's likeness has been ongoing in the entertainment industry for several years. This discussion was reignited when Deepfakes gained prevalence, after reports of celebrity's faces being used to create inappropriate videos, and demos showing how they could be used to impersonate politicians (BuzzFeed, 2018).

Due to the name Artificial Intelligence a lot of people may think that Deepfakes are intelligent, but as discussed in chapter 1, Deepfakes are a very simple application of Machine Learning. Deepfakes process a small amount of information in comparison to humans, "Its universe is limited to the training data sets, and the test data being analysed. There's nothing with which to make a relative judgment…" (Fulton, 2019). This means that they do not have the context humans have to make moral judgements and are incapable of doing so.

Additionally, there is a concept known as "instrumental neutrality" where "…technical devices are not considered the responsible moral agent…" (Gunkel, 2016). For Deepfakes this means that the tool itself is not moral or immoral, but it is up to how it is used. So, if used with consent in a non-malicious manner, it could be moral to use Deepfakes.

This is backed up in my survey data in Appendix B where 68% of respondents answered that it is ethical to use Deepfakes with consent, only 5.8% said it was never ethical. So, in the context of VFX it is ethical to use Deepfakes when we have consent from the actor. This means that there is very little difference between Deepfakes and using CG, both tools should require consent, but Deepfakes are easier to make.

Conclusion

In summary, this research report has demonstrated that in their current state Deepfakes can be used in certain applications for VFX, and further developments could give Deepfakes a wider use-case in the future. I have identified some of the technical limitations of Deepfakes and discussed how they could be solved. And I have shown where Deepfakes can and cannot fill the needs of the VFX industry.

I've found that current high-end and professional GPUs can generate high enough resolution images to have a limited application in VFX. For example, they could be used for mid to wide shots, or for use on stunt doubles, but further development would be required to replace CG. This is demonstrated in *Detective Pikachu,* where a Deepfake is used for a brief low-resolution clip.

In chapter 2 I analysed how an open source tool known as DeepFaceLab (DFL) could be used in the VFX industry. I've found that DFL has many features that would make it easy to integrate into a VFX pipeline, for example how alpha channels can be imported into compositing software such as Nuke. But some modifications would need to be made so the software could run at higher resolutions.

In chapter 3 I identified problems in my own Deepfakes using my own insight and feedback from my survey. This helped me demonstrate the requirements to create photorealistic Deepfakes, such as having enough training data from a variety of angles and lighting conditions.

Due to hardware restrictions I couldn't answer whether Deepfakes will be capable of replacing CG digital humans, but they will have a use-case in creating low budget, low resolution digital humans. I think future research should investigate how far we can push Deepfakes using cutting edge hardware such as the Nvidia DGX which can train models up to 512GB. And some research should go into finding suitable doubles who have similar head proportions. This is important since only the face is replaced, in comparison to traditional CG where the entire head can be replaced.

I think this study will be useful to those who need a guide to demonstrate where Deepfakes can be utilized in VFX and give the reader an idea of the requirements needed to create Deepfakes good enough to use in feature films.

Appendix A

Video 1

https://www.youtube.com/watch?v=-SFuqRw-On4

```
========== Model Summary ==========
==                               ==
==         Model name: SAE       ==
==                               ==
==  Current iteration: 41874     ==
==                               ==
==-------- Model Options --------==
==                               ==
==         autobackup: True      ==
==         batch_size: 6         ==
==        sort_by_yaw: False     ==
==        random_flip: False     ==
==         resolution: 256       ==
==          face_type: f         ==
==         learn_mask: True      ==
==     optimizer_mode: 3         ==
==              archi: df        ==
==            ae_dims: 300       ==
==          e_ch_dims: 32        ==
==          d_ch_dims: 16        ==
== multiscale_decoder: True      ==
==         ca_weights: True      ==
==         pixel_loss: True      ==
==   face_style_power: 0.1       ==
==     bg_style_power: 4.0       ==
==    apply_random_ct: True      ==
==           clipgrad: True      ==
==                               ==
==--------- Running On ----------==
==                               ==
==       Device index: 0         ==
==               Name: Tesla K80 ==
==               VRAM: 11.00GB   ==
==                               ==
===================================
```

Video 2

https://www.youtube.com/watch?v=kDwxFkg1lvU

```
============ Model Summary ============
==                                   ==
==        Model name: SAE            ==
==                                   ==
== Current iteration: 85725          ==
==                                   ==
==----------- Model Options ---------==
==                                   ==
==        autobackup: True           ==
==       sort_by_yaw: False          ==
==       random_flip: False          ==
==        resolution: 256            ==
==         face_type: f              ==
==        learn_mask: True           ==
==    optimizer_mode: 3              ==
==             archi: df             ==
==           ae_dims: 360            ==
==         e_ch_dims: 32             ==
==         d_ch_dims: 16             ==
==        ca_weights: False          ==
==        pixel_loss: True           ==
==  face_style_power: 0.1            ==
==    bg_style_power: 4.0            ==
==   apply_random_ct: True           ==
==          clipgrad: True           ==
==        batch_size: 4              ==
==                                   ==
==------------ Running On -----------==
==                                   ==
==      Device index: 0              ==
==              Name: GeForce GTX 1070 ==
==              VRAM: 8.00GB         ==
==                                   ==
=======================================
```

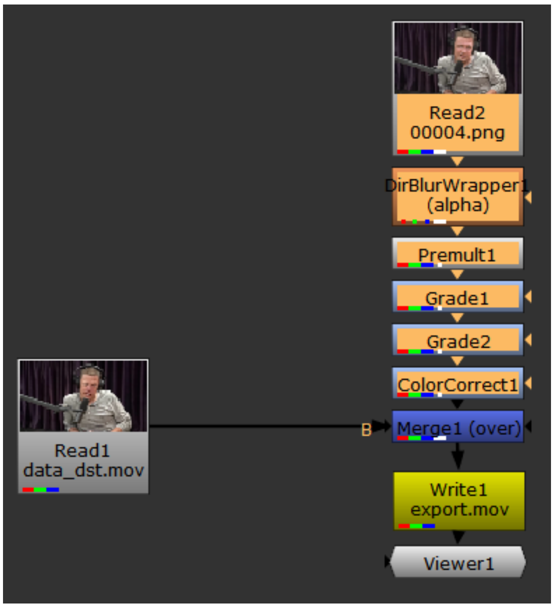
Video 3

https://www.youtube.com/watch?v=yNd7LF3dN98

```
============ Model Summary =============
==                                    ==
==         Model name: SAEHD          ==
==                                    ==
==  Current iteration: 70337          ==
==                                    ==
==----------- Model Options -----------==
==                                    ==
==         autobackup: True           ==
==        sort_by_yaw: False          ==
==        random_flip: False          ==
==         resolution: 128            ==
==          face_type: f              ==
==         learn_mask: True           ==
==     optimizer_mode: 2              ==
==              archi: df             ==
==            ae_dims: 512            ==
==         ed_ch_dims: 21             ==
==   face_style_power: 10.0           ==
==     bg_style_power: 10.0           ==
==    apply_random_ct: True           ==
== true_face_training: True           ==
==           clipgrad: False          ==
==         batch_size: 6              ==
==                                    ==
==------------- Running On ------------==
==                                    ==
==       Device index: 0              ==
==               Name: GeForce GTX 1070 ==
==               VRAM: 8.00GB         ==
==                                    ==
========================================
```

Video 4
https://www.youtube.com/watch?v=emfq0T84A1E

```
========= Model Summary =========
==                             ==
==         Model name: SAEHD   ==
==                             ==
==  Current iteration: 57689   ==
==                             ==
==-------- Model Options --------==
==                             ==
==        sort_by_yaw: False   ==
==        random_flip: True    ==
==         resolution: 144     ==
==          face_type: f       ==
==         learn_mask: True    ==
==     optimizer_mode: 3       ==
==              archi: df      ==
==            ae_dims: 512     ==
==         ed_ch_dims: 21      ==
==        random_warp: True    ==
== true_face_training: True    ==
==   face_style_power: 0.0     ==
==     bg_style_power: 0.1     ==
==            ct_mode: rct     ==
==           clipgrad: False   ==
==         batch_size: 8       ==
==                             ==
==---------- Running On ----------==
==                             ==
==       Device index: 0       ==
==               Name: Tesla K80 ==
==               VRAM: 11.00GB ==
==                             ==
=================================
```

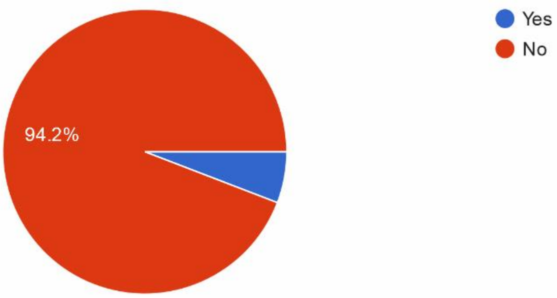


Appendix B

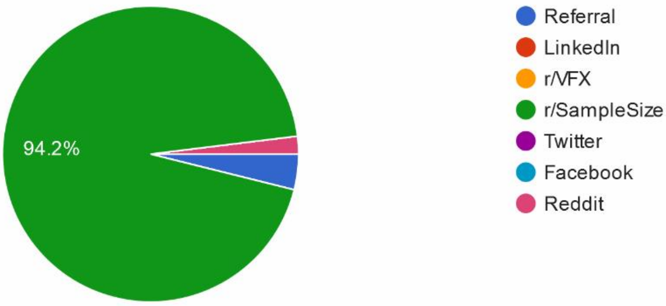
Survey URL: https://docs.google.com/forms/d/e/1FAIpQLSd-EYGr1x8fvAhahpyRPJKaBsotDaWUtabsN6zn6AXQ4TH-pw

Spreadsheet: https://docs.google.com/spreadsheets/d/e/2PACX-1vTixQPHdgSBFrfQmK_cvuHjjOi7hn4vYW83Qr6HruRQe2iFmoaQf4lGn702TIM01WmE8B0C3xTgugrL/pubhtml

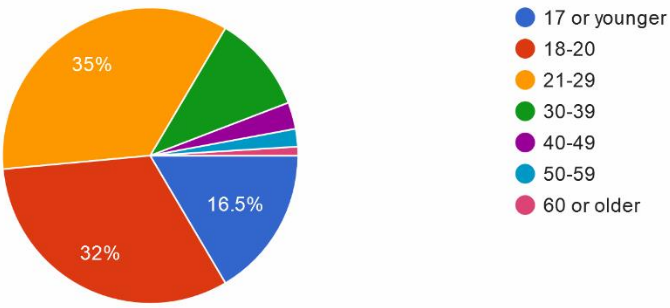
Are you involved in the VFX industry? (includes students and enthusiasts)
103 responses



Where did you find this survey?
103 responses



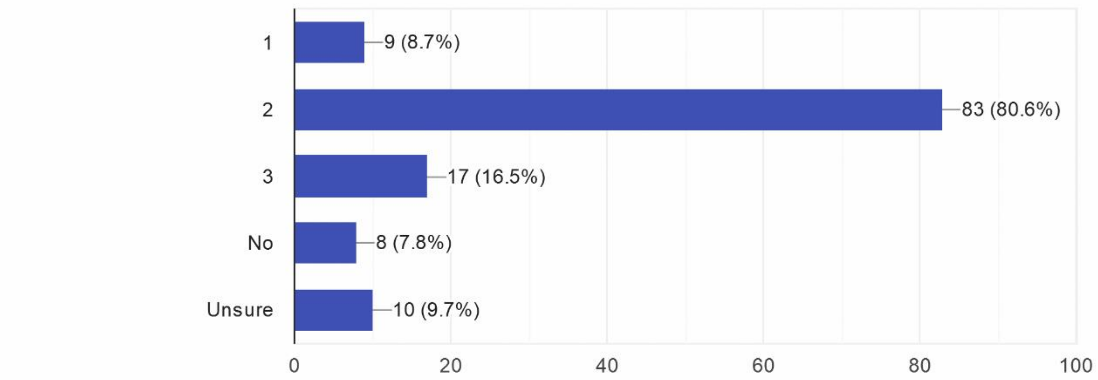
Age
103 responses



Video 1
https://www.youtube.com/watch?v=Flrx1CrPjck

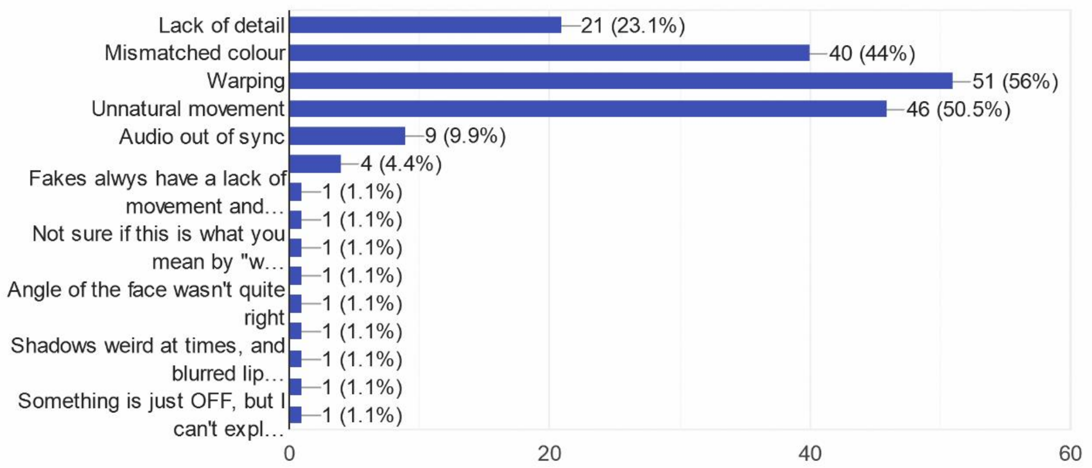
Were any of these videos fakes? If so which?

103 responses



If you noticed a fake why did you notice it?

91 responses

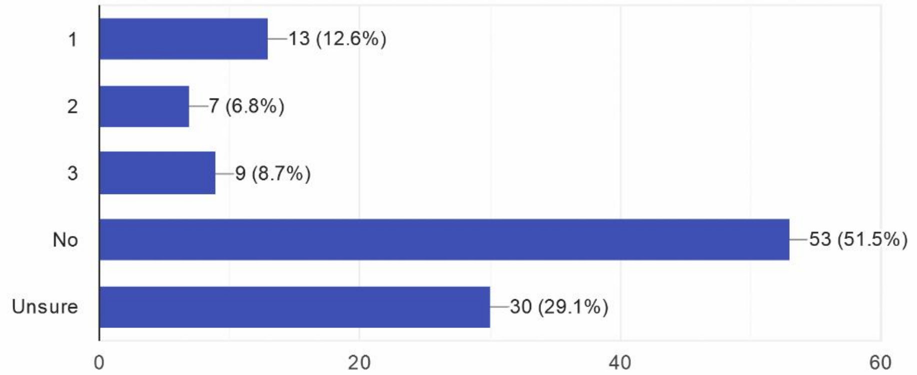


Video 2

https://www.youtube.com/watch?v=L6NpYMt-CzM

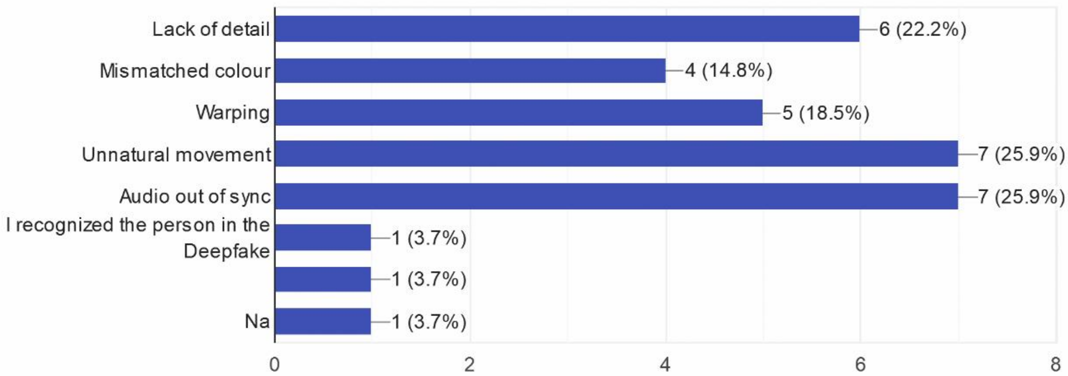
Were any of these videos fakes? If so which?

103 responses



If you noticed a fake why did you notice it?

27 responses

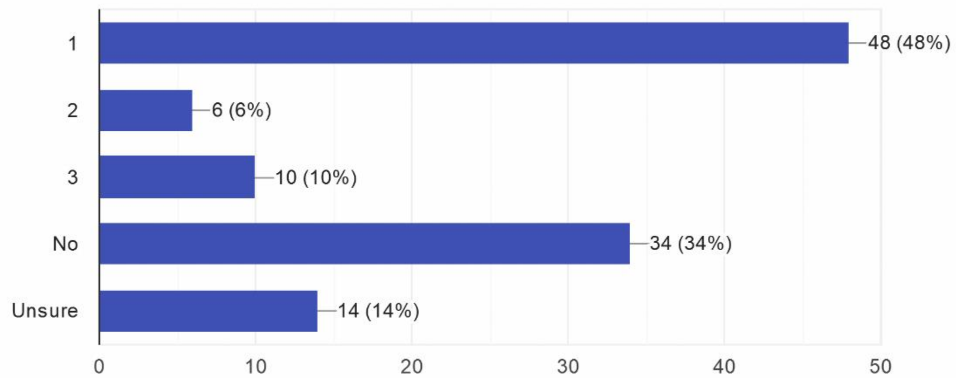


Video 3

https://www.youtube.com/watch?v=nEVoGhW9Qek

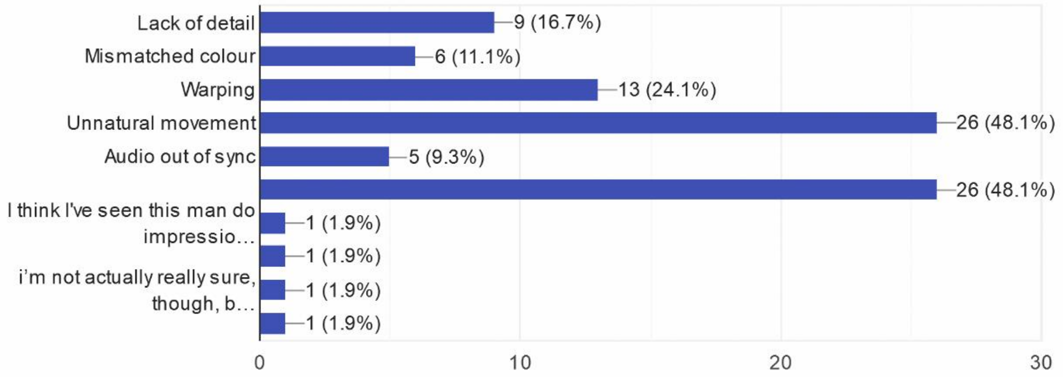
Were any of these videos fakes? If so which?

100 responses



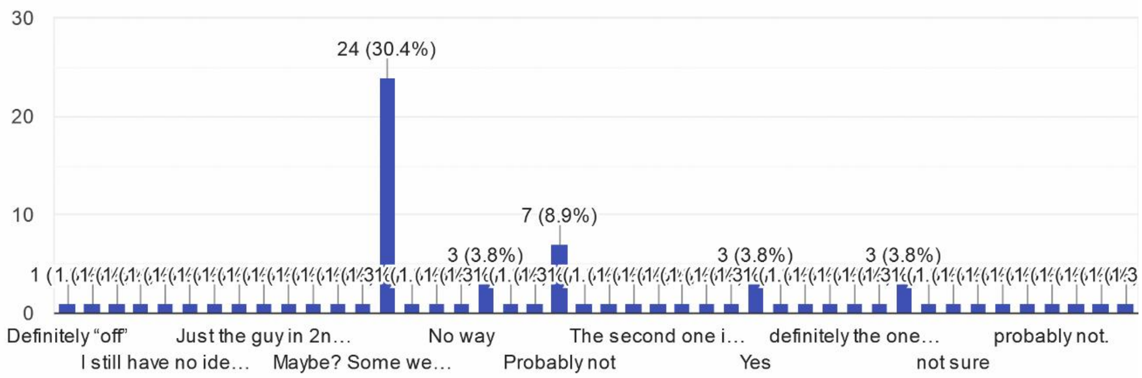
If you noticed a fake why did you notice it?

54 responses



Would you have noticed any fakes if you weren't asked to look for them?

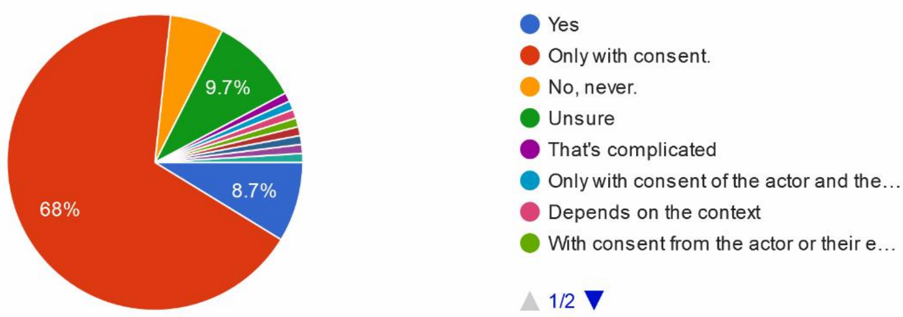
79 responses



Ethics

Is it ethical to use Deepfakes to recreate an actors face?

103 responses



Bibliography

3Blue1Brown, 2018. *Neural Networks.* [Online]
Available at: https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi
[Accessed 1 10 2019].

Apple, 2019. *Apple unveils powerful,.* [Online]
Available at: https://www.apple.com/uk/newsroom/2019/06/apple-unveils-powerful-all-new-mac-pro-and-groundbreaking-pro-display-xdr/
[Accessed 5 11 2019].

Bradshaw, T., 2019. *Deepfakes: Hollywood's quest to create the perfect digital human,* s.l.: Financial Times.

BuzzFeed, 2018. *You Won't Believe What Obama Says In This Video!* 😉*.* [Online]
Available at: https://www.youtube.com/watch?v=cQ54GDm1eL0
[Accessed 8 May 2019].

CyberDainz, 2019. *Introducing the new SAEHD model..* [Online]
Available at:
https://www.reddit.com/r/SFWdeepfakes/comments/dfgv68/introducing_the_new_saehd_model/
[Accessed 23 October 2019].

Dmytro Korolov, D. R. J.-C. B. R. Z. R. P. J. K. D. L., 2018. Future of artificial intelligence and deep learning tools for VFX. *ACM SIGGRAPH 2018 Panels.*

Fauer, J., 2014. *Film & Digital Times: Alexa 65 Special.* [Online]
Available at: http://www.fdtimes.com/pdfs/issues/65FDTimes-ALEXA65v6.0-150.pdf
[Accessed 5 11 2019].

Fulton, S., 2019. *What is bias in AI really, and why can't AI neutralize it?.* [Online]
Available at: https://www.zdnet.com/article/what-is-bias-in-ai-really-and-why-cant-ai-neutralize-it/
[Accessed 12 October 2019].

Gerrish, S., 2018. *How Smart Machines Think.* The MIT Press: Cambridge, Massachusetts.

Glassner, A., 2018. *Deep Learning: A Crash Course.* [Online]
Available at: https://www.youtube.com/watch?v=r0Ogt-q956I
[Accessed 5 May 2019].

Glassner, A., 2019. *A Crash Course in Deep Learning.* Stuttgart, FMX.

Güler, R. A., Neverova, N. & Kokkinos, I., 2018. *DensePose: Dense Human Pose Estimation In The Wild.* [Online]
Available at:
https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=edsarx&AN=edsarx.1802.00434&site=eds-live&custid=ns195502
[Accessed 25 March 2019].

Gunkel, D. J., 2016. Paradigm Shift: Media Ethics in the Age of Intelligent Machines. *Controversies in Digital Ethics,* pp. 233, 247.

Hsu, J., 2012. *Why "Uncanny Valley" Human Look-Alikes Put Us on Edge.* [Online]
Available at: https://www.scientificamerican.com/article/why-uncanny-valley-human-look-alikes-

put-us-on-edge/
[Accessed November 2 2019].

iperov, 2019. *DeepFaceLab.* [Online]
Available at: https://github.com/iperov/DeepFaceLab
[Accessed 5 2019].

iperov, 2019. *manual_en_google_translated.pdf.* [Online]
Available at:
https://github.com/iperov/DeepFaceLab/blob/master/doc/manual_en_google_translated.pdf
[Accessed 5 2019].

Korshunov, P. a. M. S., 2019. *DeepFakes: a New Threat to Face Recognition? Assessment and Detection.* [Online]
Available at: https://arxiv.org/abs/1812.08685
[Accessed 5 April 2019].

Köse, U., May 2018. Are We Safe Enough in the Future of Artificial Intelligence? A Discussion on Machine Ethics and Artificial Intelligence Safety.. *BRAIN: Broad Research in Artificial Intelligence & Neuroscience,* pp. 184-197.

Livingstone, S. R., 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English.. *PLoS ONE,* 5 April.p. 13.

Li, Y. & Lyu, S., 2018. *Exposing DeepFake Videos By Detecting Face Warping Artifacts.* [Online]
Available at:
https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=edsarx&AN=edsarx.1811.00656&site=eds-live&custid=ns195502
[Accessed 25 March 2019].

Masahiro Mori, K. F. M. N. K., 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine,* Volume 19, pp. 98-100.

mc.ai, 2019. *Auto-encoder in biology.* [Online]
Available at: https://mc.ai/auto-encoder-in-biology/
[Accessed 5 11 2019].

Mire, A., Dhok, S. B., Mistry, N. J. & Porey, P. D., 2013. Resampling Detection in Digital Images: A Survey. *International Journal of Computer Applications,* Volume 84, pp. 24-29.

Oberoi, G., 2018. *Exploring Deepfakes.* [Online]
Available at: https://www.kdnuggets.com/2018/03/exploring-deepfakes.html
[Accessed 18 March 2019].

Pokémon Detective Pikachu. 2019. [Film] Directed by Rob Letterman. United States, Japan: The Pokémon Company.

Price, A., 2018. *The Next Leap: How A.I. will change the 3D industry.* [Online]
Available at: https://www.youtube.com/watch?v=FlgLxSLsYWQ
[Accessed 18 March 2019].

Stewart, M., 2019. *Comprehensive Introduction to Autoencoders.* [Online]
Available at: https://towardsdatascience.com/generating-images-with-autoencoders-77fd3a8dd368
[Accessed 5 11 2019].

Tinwell, A., 2014. *The Uncanny Valley in Games and Animation.* 1st ed. s.l.:CRC Press .

Umé, C., 2019. *DAVID HASSELHOF DEEPFAKE DE-AGING TURNING BACK THE TIME.* [Online]
Available at: https://www.youtube.com/watch?v=jYY0LAT1YRQ
[Accessed 1 November 2019].

Xu, N., Price, B., Cohen, S. & Huang, T., 2017. *Deep Image Matting.* [Online]
Available at:
https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=edsarx&AN=edsarx.1703.03872&site=eds-live&custid=ns195502
[Accessed 20 March 2019].

Zucconi, A., 2018. *Understanding the Technology Behind DeepFakes.* [Online]
Available at: https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes/
[Accessed 18 March 2019].