Convergence: The International Journal of Research into New Media Technologies
'The Digital Face and Deepfakes on Screen'

The digital resurrection of Margaret Thatcher: creative, technological and legal dilemmas in the use of deepfakes in screen drama.

INTRODUCTION

Deepfakes have been used by practitioners since 2017 as a cheap and rapid means of face replacement in digital video. Deployed with both playful and malign purposes, as well as by artist-activists as political provocations, the practice has paraded a spectacle of technological subculture, a nose-snubbing to the high-end Digital Visual Effects (DVFX) face replacements perfected by Disney Studios for major Hollywood blockbusters. This oppositional subculture initially reveled in its low-grade achievements, with early deepfakes (DFs) produced at very low resolution, typically no more than 294x500 pixels. While recent research has strived to upgrade resolution (Naruniec et al 2020), the world of deepfakes continues to exist almost exclusively within the confines of a handheld screen culture in which such low-quality images are acceptable. This paper examines the potential development of the deepfakes application of machine learning into mainstream screen production. One of the first such examples of this was seen in December 2020, when the UK broadcaster, Channel Four, aired an *Alternative Christmas Message* that used deepfake processes to create a synthetized Queen Elizabeth II to address her nation[i]. The program's director, William Bartlett, emphasized that the purpose was to illustrate the unreliability of the moving image in the era of deepfakes, although audience reactions pointing to the unbelievability of the depiction of the monarch[ii] undermined the sense of threat. Bartlett was working with the full resources of one of the biggest DVFX postproduction houses, Framestore, so this deepfake was still embedded in the elite world of high-budget filmmaking. In this article, we predict a near-future in which open source machine learning has democratized access to face replacement in High Definition digital video; we discuss questions around the technological routes to this goal of high quality DFs in fiction production, as well as the major ethical and legal issues that creative practitioners will face when choosing this tool. Our work focusses on the creation of synthetized screen characters based on deceased celebrities, developing from an ongoing interdisciplinary practice research project, *Virtual Maggie*, that explores whether Margaret Thatcher could be digitally resurrected, using machine

learning, to play herself in a new period drama[iii]. The current article is both a report on the research findings of *Virtual Maggie*, as well as a widening consideration of the issues concerned, and is structured around the three themes of that investigation:

- *Creative screen practice:* how are the processes, relationships and responsibilities of the filmmaker changed when applying deepfakes to the construction of screen characters?
- *Technological practice:* what are the methodological choices available in using machine learning for character face replacement using historical figures and what level of visual quality can be achieved?
- *Legal context:* to what extent is UK and international law prepared for the exploitation of individuals' images after their death and what legal considerations should be taken by practitioners when creating deepfakes of major historical figures?

Consistent with the interdisciplinary nature of the *Virtual Maggie* project, this article is designed to have relevance within three traditions of academic study: film studies/practice, computer science, and law. The three sections of the article are expressed in the language of these disciplines, before we bring together the findings of the research to discuss shared conclusions.

CREATIVE SCREEN PRACTICE

Theories of screen performance

The digital replacement of an actor's face further complicates our understanding of the nature of screen performance, adding a layer of technological intervention to a process that has long been understood as a composite of creative inputs. James Naremore (1986) discussed the 'expressive coherence' of multiple elements of a single complete film performance. These could be different layers of performance achieved by actors while creating their screen characters, but the term also describes the fragmented and recombined elements of performance due to the processes of filmmaking. Naremore also stressed that a screen performance may involve the labor of several individuals, pointing out that 'movies are the only medium in which several actors are *typically* used to play one role' (1986, 50), referring to the work of stunt performers, voice actors and body doubles that supplement the actor's performance. In addition to this, the labor of a film's editor, sound editor and colorist have always added subtle contributions to the creation of a film character presented to the audience. This does not negate the primary role of the actor, rather it emphasizes the breadth of the creative effort required to generate screen performance. The recent addition of digital face replacement adds significant new human and technological inputs to the collective process of building a film character. For Lisa Bode, this

represents a conceptual challenge: ‘we need to examine what is actually achieved when performance, technology, special and visual effects, and animation work together both on and behind the screen’ (2017, 11). As a historian of cinema, Bode’s work (2007, 2010) establishes a context for current debates about digital resurrection in the context of deepfakes, providing an early definition of what is now becoming known as ‘performance synthetization’ - the digital manipulation of a performance or a performer’s likeness (Pavis 2020).

Virtual Maggie

A significant quality of the *Virtual Maggie* practice research project is its reflection of real-world film industry creative processes. The origins of this project stem from a feature screenplay, *Rebel Bus*, an as-yet unproduced drama set in South Wales and Northern Ireland in 1989. The narrative includes a small role for Margaret Thatcher, in which the prime minister responds to the disruption of her handling of the ‘The Troubles’ caused by an unlikely group of Welsh football fans who follow their team to a sporting fixture in a war zone – the IRA-controlled Bogside of Derry. In the film industry, the ‘development’ stage of preproduction includes the drafting of the screenplay, followed by important early work on casting, which will have a major influence over the successful financing of the picture. In the case of *Rebel Bus*, the part of Margaret Thatcher was too small to be attractive to actors who had successfully played her on screen in earlier movies, such as Meryl Streep (*The Iron Lady*, Phyllida Lloyd, 2011). The filmmakers then considered the option of digital resurrection: instead of asking the audience to believe the interpretation of Thatcher by an actress, technology could allow them to build a hybrid screen character, *Virtual Maggie*, using the body of an actor combined with a digitally re-rendered face of Margaret Thatcher herself. Hollywood’s high-budget DVFX approach to performance synthetization was clearly not an option for an independent film production in its early concept stages; the only available route to the goal of creating *Virtual Maggie* was to adopt a machine learning, or Deepfake approach.

Such decision-making remains innovative in the screen industries, in which there is still little adoption by mainstream producers of digital face replacement as a creative tool. The choice to pursue a machine learning route to creating *Virtual Maggie* was optimistic, requiring an extended process of technological research. However, this practice research project foresees that within a few years of improvement of machine learning, such a scenario of decision-making may take place with frequency in film and television production companies of all sizes. Machine learning represents a credible future for creative decision-makers across the screen industries, giving the *Virtual Maggie* research project timeliness and urgency.

Limitations of technology

The discussion of technological breakthroughs is frequently framed by an ideology of potential, a belief that the development of new techniques will open up limitless opportunities for creativity or productivity. The reality is that each new process generates its own limitations, requiring its adopters to conform to the characteristics dictated to them by the structure and design of the new technology. The *Virtual Maggie* project enabled its researchers to study the limitations of machine learning-based face replacement at its current stage of development. Two key issues will be developed, illuminating how machine learning creates considerable challenges for the creative process of filmmaking.

A common characteristic of the most well-known deepfakes is how the camera is used. A recurring pattern emerges: the camera is locked-off, usually framing its subject in a mid-shot or loose MCU. Furthermore, a typical performance style is repeated: the actor/subject is seated or standing still.

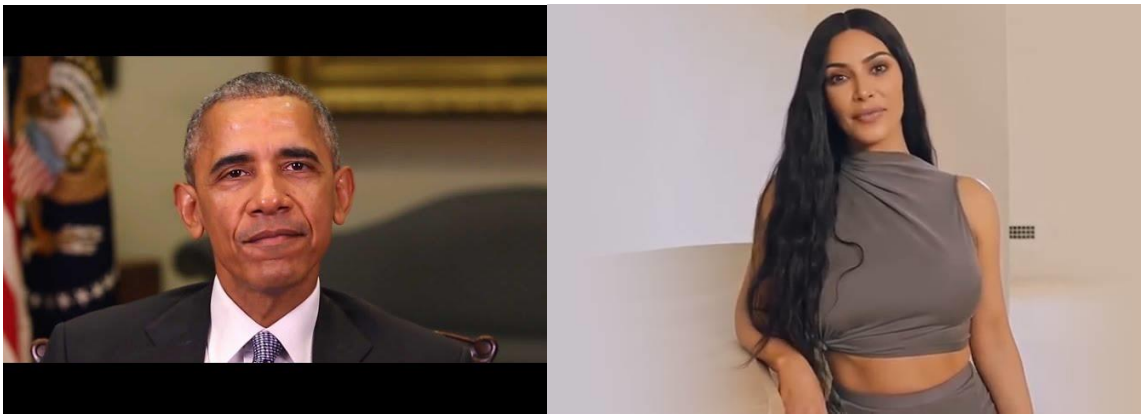

Figure 1: Locked-down screen performance: screengrabs from Jordan Peele's DF of Barack Obama; DF of Kim Kardashian by Bill Posters

The strategy being used here is framed to accommodate a weakness of the use of machine learning processes in performance synthetization. Any movement of the face in relation to the camera, either by the performer or through camera movement, creates major additional requirements in the processing of the composite digital image. Many of the high-profile deepfakes, such as those in Figure 1, have been based on the manipulation of just the mouth and lips of the subject; without the need to engineer fake head movements, the AI task becomes relatively straightforward. In comparison, the task of digital face replacement in a normal film drama, with changing shot sizes, expressive acting, and camera movements, becomes a much more significant challenge. Scale of face is also important. Deepfakes have,

from their origin, been produced at low resolution, typically no more than 294x500 pixels. Despite the recent research seeking to upscale this resolution (Naruniec et al 2020), in any digital face replacement the size of the face within the image will have a major bearing on the efficacy of the AI process. In Figure 1, we can see that the framing adopted by Bill Posters when working on his deepfake of Kim Kardashian was a shrewder choice than that of the Jordan Peele deepfake of Barack Obama: whereas Kardashian's face occupies less than a quarter of the vertical space of the frame, Obama's covers more than half.

During the practice research of the first stage of the *Virtual Maggie* project, the filming of scenes in studio and on location, the impact of this limitation of the deepfake technology became apparent. The director and cinematographer were constrained while constructing shots of Margaret Thatcher. Advised on set by the project's computer scientist, continuous attention to the scale of the actress's face in the frame was necessary. Certain shot sizes, such as Close-up and Medium Close-up, were abandoned because of the challenges that this would create to the face replacement process.

Figure 2: On set during the Virtual Maggie shoot: Medium Shot of actress Ros Adler

A second significant limitation of the machine learning process also impacted on decision-making. Attentive followers of deepfakes will have noticed that the eyeline of synthetized characters is always very close to the lens. In most non-fiction deepfakes, the subject addresses the camera directly (cf Fig.1); in the small number of fictitious scenes that include deepfakes, the character undergoing digital face replacement is almost always facing towards the camera. This is because the machine learning process finds it particularly difficult to successfully replace a face seen in a side angle, a limitation that has profound implications for

the conventions of cinema. We will take an example from one scene in the *Virtual Maggie* sequence, in which the Prime Minister and her Secretary of State for Northern Ireland, Tom King, are travelling in a car. The culture of cinema has developed a portfolio of customary shot choices for travelling car scenes, and many of these are derived from the practical problems of positioning a film camera in, or on, a moving vehicle. Frontal angles shot using a camera mount on the car bonnet or a low-loader camera vehicle are supplemented by over-the-shoulder (OTS) profile shots of the actors, filmed either from within the car or using camera mounts on the doors. The *Virtual Maggie* scene begins with a standard loose frontal two-shot of Margaret Thatcher and Tom King sitting side-by-side. The crew then set up and filmed two complementary OTS side angles of the characters (Fig 3), following a typical choice of shots coherent with cinematic tradition.

Figure 3: Interior car scene - preparing side angle shot for Virtual Maggie

Following a consultation with the project's Co-Investigator responsible for carrying out the deepfakes processing in the project, it was decided that such side angles could be extremely difficult to integrate into the deepfake workflow: two supplementary shots were added to the schedule, Medium Shots of each character filmed from within the car that specifically avoided the side angle position, with both Margaret Thatcher and Tom King's eyelines close to camera. This was essential in order to give flexibility in postproduction: if our machine learning process were to be incapable of digitally replacing actress Ros Adler's face in the OTS profile of Maggie, we could resort to re-editing the scene using just the supplementary shots.

The practice research of the *Virtual Maggie* project has illuminated key alterations to the portfolio of creative choices available to a filmmaker when using deepfake technology, specifically the shots sizes and camera angles that are possible. Other early practitioners, such

as William Bartlett in his *Alternative Christmas Message* (2020), demonstrate an awareness of these problems in their language of camera framing. Bartlett's Elizabeth II is framed frontally throughout, including in her desktop dance sequence (Fig 4).

Figure 4: *Alternative Christmas Message* (William Bartlett, 2020)

If deepfakes become a mainstream creative choice of film and television producers in the next few years, a potential impact on screen culture may arise. The limitations of the technology will influence creative choice, significantly limiting how the digital film camera captures action. Camera movement may also become restricted in dramas using deepfakes: Bartlett's opening shot, a crabbing movement infront of the Queen's desk, is one of the weakest in terms of the believability of his deepfake project, illuminating further deficiencies of the technology that future filmmakers may seek to avoid by adopting static camera positions during deepfake sequences.

The issue of how technological requirements of the deepfake process cause creative constraints in cinematography fits into a pattern across the history of cinema, in which breakthroughs in technology impact other aspects of the creative process, most frequently the work of actors. In the early 'talkies' era, the practice of hiding the microphone behind a prop limited the blocking of actors, who could not deliver lines at any distance from this part of the set. The telescopic boom was quickly developed to overcome this problem. More recently, the common use of greenscreen cinematography forces the actor to perform in isolation instead of in an ensemble, an experience that nearly led to Sir Ian McKellen abandoning his career when he played in *The Hobbit: An Unexpected Journey* (Peter Jackson, 2012)[iv]. Current deepfakes technology should be seen as the beginning of a process in which the screen industries first adopt new processes, then confront their limitations, before adapting production practices in order to address the new constraints posed by advances in screen technologies.

A further insight from the film shoot stage of the *Virtual Maggie* project has been the importance of the director-technologist creative relationship. Central to the successful construction of digitally synthetized characters is the level of creative and technical understanding between these two key individuals. Throughout the film's preproduction and during the shoot, an ongoing dialogue about the potentials and the limitations of machine learning contributes to each of the decisions made. This collaboration then flows into the second stage of producing a deepfake screen character - the machine learning process that begins when the editing of scenes is complete.

TECHNOLOGICAL PRACTICE

Challenges of creating deepfakes of historical figures

The deep fake methods used to synthetize the appearance of one person to another are largely based on the ability of deep neural networks to learn a representation of multiple facial poses of one face, and transfer that pose to a second face. Underpinning this is a reliance on a large volume of exemplar material which is required in order to successfully train the neural networks to accurately carry out the task. While there is a large amount of video content for contemporary actors and personalities in the public domain, this is not the case for the domain of creating deep fakes of historical figures. Note, we are limiting this to figures of whom accurate imagery exists, for example photography and video; the extension to other sources such as paintings is outside the domain considered in this work. The first limitation encountered is the small volume of information: while there exists video footage of many historical figures, this is typically less than for contemporary figures. Secondly, this footage is likely to be significantly lower quality due to technological limitations at the time, degradation of the content and the digitization process. Thirdly, much footage is likely to be black and white which again makes it challenging for use in modern color productions.

In this section, we create a framework which can solve these technical challenges whilst also considering ease of use by end users. One insight is that many of these challenges have been tackled by the machine learning, computer vision and computer graphics field, but are yet to be fused into a pipeline for historical facial replacement. This framework has been implemented into a tool which is designed to produce frames for the *Virtual Maggie* project. We designed this tool considering several factors: the process should lead to plausible face swap results, the

limitations outlined above should be circumvented, and minimal user interaction should be required to generate the final imagery.

Related Work

There are multiple approaches for swapping faces, from traditional approaches such as warping a source face to a target face considering 3D geometry (Blanz, et al., 2004) to models using deep learning (Naruniec, et al., 2020). We focus on deep learning approaches for face swapping[v] as these form the current state-of-the-art and more information can be found in the survey by Nguyen et al. (Nguyen, et al., 2019).

Early approaches to generating deepfakes were proposed outside academia (DeepFakes, 2020). These used an encoder network combined with two decoder networks; the shared encoder network encodes source and target faces into a shared latent space (Liu, et al., 2017), and the two decoder networks reconstruct the source and target images from the latent representation. The approach by Naruniec et al. (2020) both generalized the number of outputs, and utilized high resolution inputs and outputs to lead to film quality face swapping. However, all these approaches reply on a large volume of source and target data to produce viable results. In the context of this work, it is expected that there is a large volume of target actor material, but a limited amount of source material which leads to these encoder-decoder approaches being unsuitable for the historical deep fake context.

Alternative approaches for face replacement which do not rely on large volumes of training data are to use Generative Adversarial Networks (GAN) (Goodfellow, et al., 2014) to replace faces, or to replace features in one image with those in another. Nirkin et al. (2019) proposed a GAN to fill-in segmented regions of a target image with a source image. While this can use a relatively small amount of data, the results do not generate image quality suitable for film usage. Lathuiliere et al. (2020) propose an approach which can directly swap parts between two images while retaining natural appearance. This approach learns how to segment faces into constituent parts and enables these constituent pieces to be transferred to another frame through estimating optical flow (Horn and Schunck, 1981). We base our work on this approach as it is suitable for our requirements: it requires one source image, is capable of being used with high resolution images, and is relatively robust to changes in facial pose.

Framework

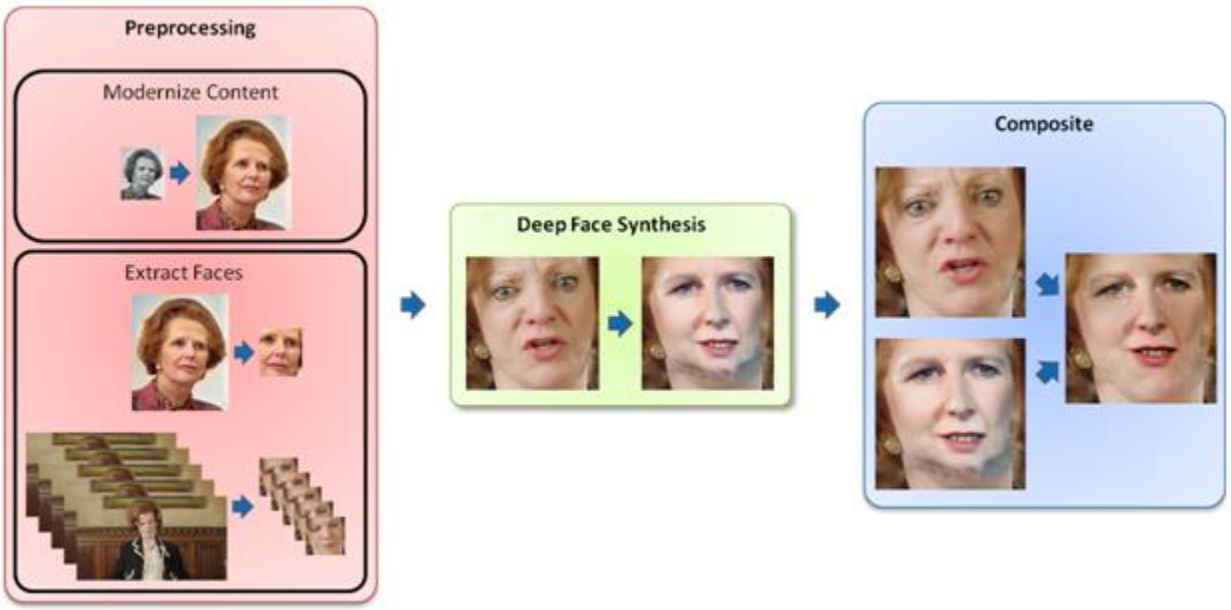

Figure 5: The flow of data in our proposed framework. Initial data sources include a historical image, which can be modernized, and a series of frames. Faces are then automatically extracted from both these sources. The next stage of the pipeline is the face swap component which swaps the extracted faces from each frame of the animation and the modernized historical image. Finally, these are composited such that the swapped face has the same lighting and skin appearance as the captured footage.

We propose a three-step framework for face swapping for historical images. The first step is *Preprocessing* where faces are extracted from the historical source image and the target frames, and the historical image is modernized if required. The second step is to perform *deep face synthesis* using a deep learning model. This is a 'black box' in our framework meaning that as future methods are developed, they may be swapped for the existing model if required. The final aspect is *compositing* the swapped face back into the target frames, while preserving the original color and illumination information. We propose that this framework should consist of several 'black box' modules which can be swapped for other modules to keep pace with developments in the machine learning and computer graphics fields. The following sections discuss these aspects in more detail and an overview is shown in Figure 5.

Preprocessing

Preprocessing of the data serves two purposes. The first is to modernize the source image if required, and the second is to extract faces. The historical source image needs to match the desired color depth of the target frames. For example, if the target footage is shot in color, then the source image also needs to be in color. However, due to camera limitations at the time of capture, many source images are black and white. These therefore need to be converted to color. Secondly, the resolution of the source imagery needs to be sufficient to match the requirements of use in film. If imagery is scanned from physical film, then this is likely to produce source images at the desired resolution; however, this cannot be guaranteed for existing digital content. Finally, much existing historical imagery and footage has degraded quality compared to modern footage. This is due to damage to the source film over time, or low quality of the capture devices at the time of recording.

Modernization

Colorization of black and white images can be tracked through deep learning methods. Several approaches have been proposed, ranging from a fully connected network combined with filtering (Cheng, et al., 2015), to convolutional neural network approaches (Zhang, et al., 2016), GANs (Nazeri, et al., 2018) and networks designed for historical footage (Iizuka and Simo-Serra, 2019; Antic, 2020).

Upscaling of imagery refers to starting with a low-resolution image and generating a higher resolution output. This has been tackled using many different approaches, from traditional computer vision and image processing methods such as bicubic upsampling to CNNs (Dong, et al., 2014) and GANs (Dong, et al., 2015). For more information about these approaches, see the review by Anwar et al. (Anwar, et al., 2020).

Degradation removal was also tackled by Iizuka and Simo-Serra (2019). This work synthesised film deterioration effects and applied it to existing footage, then trained a network to remove this degradation. Other approaches can detect and remove specific artifacts for historical footage (Helm and Kampel, 2020) or noise in images (Yuzhi, et al., 2020).

These steps only need to be applied if the historical image requires modernisation, and steps may be omitted: for example, if a source image is in color but low resolution, then only the upscaling step is required.

Face Extraction

In order for the deep face synthesis stage to swap faces, it is either desirable, or sometimes essential, that the input to the deep learning system only contains faces. Therefore, once the source and target images are obtained, the second step of preprocessing is to extract and crop faces from the imagery. The historical source image is required to contain the image of the historical figure, and no other information, whereas the target images may contain other actors. This stage of preprocessing automatically detects faces in both the source and target images, and resizes them to the required resolution for input into the deep learning system.

Face detection can be performed using traditional image processing techniques, such as extracting keypoints corresponding to facial features (Viola and Jones, 2001; Wilson and Fernandez, 2006) and creating a bounding box around these features, or more modern deployed learning face detection techniques can be used. Regardless of the technique used, bounding boxes around faces are computed for each image and then resized to the required input resolution for the deep face synthesis network.

This however leads to two issues in an automated system. The first is how to deal with the situation that the target frames contain multiple actors, and therefore multiple faces. The second is how to handle temporal stability of the detected bounding boxes. The first issue can be dealt with by first detecting all faces in an image, then comparing face statistics to those of the face of the actor whose face should be swapped. The bounding box corresponding to the statistics which most closely match those of the target actor is kept, and the remaining are discarded. The approach we used in our system was based on Kazemi and Sullivan (2014). The second issue was discussed in Naruniec, et al. (2020) and we follow a similar approach by randomly cropping a larger region around the initially detected face and re-running the face detection algorithm. The average of the resulting bounding boxes, which is itself another bounding box, provides a temporally stable estimate of the position of the face in each frame. We also found the parameters described in Naruniec, et al. (2020) worked well for our imagery.

Deep Face Synthesis

The second stage of the framework is the deep face synthesis aspect which swaps the faces in the source and target crops. As our framework has to work in the situation of limited historical data, use of deep learning systems which consist of an encoder and decoder may not be feasible as there may be a lack of the data required for training the decoder. Therefore, we propose this step should operate using the single source historical image. Fortunately, this is achievable using methods based on segmentation and optical flow prediction. We use the approach proposed by Lathuiliere et al. (2020) which both segments individual features, for example eyes, cheeks, jaw, from both the source and target image, then based on the estimation of optical flow, deforms each segmented region in the source image to match the corresponding segmented region in the target image. The results of this can be seen in the middle image in Figure 5.

Compositing

The result of the *deep face synthesis* stage creates a face with appropriate pose, but preserves the skin tone and lighting from the source image, and may contain artifacts on the boundaries of segments. The incorrect skin tone, lighting, and remaining artifacts are corrected in the compositing stage. Boundary artifacts are removed through replacing the actor's face in a masked region created from keypoints created during the face extraction stage to ensure the central area of the face is replaced, rather than the boundaries which are likely to contain artifacts.

In order to correct for skin tone and lighting information, we decompose both the swapped face and the original actor's face into a Laplacian pyramid. Each level of the pyramid contains progressively lower frequency details, and the lower levels generally encode coarse skin tone and lighting information. Similar to Thies, et al. (2015) and Naruniec, et al. (2020), we use the skin tone and lighting information from the actor's face encoded at the lower levels (we use the first two levels) then reconstruct from the remaining levels in the Laplacian pyramid of the swapped face. This preserves skin tone and lighting across the face, while simultaneously preserving the appearance of the swapped face containing the historical figure.

Our Implementation

Figure 6: Screenshot of the user interface. The optional modernisation functionality is enabled by the three buttons, and the ‘Run’ button automatically runs the remaining stages of the framework.

We implemented a prototype of the framework as described above. For ease of use, we developed a simple user interface to allow the process to occur with a minimal amount of user input (see Figure 6). Initially, a source historical image is loaded (Figure 6, top) and the user is presented with options to run any of the modernization functionality if required. Then target frames are loaded, and the user can select the face of the actor whose face will be replaced in the first frame (to build statistics for the face extraction stage). Finally, once an output folder is specified, the remaining steps of the framework are run automatically when the "Run!" button is pressed. On average this takes around 10 seconds to process each frame on a laptop with a Nvidia 1050 GPU, although this average takes into account that each stage of the framework is currently run to completion before moving to the next stage of the framework, for example the face extraction stage is run for all frames before moving on to the deep face synthesis stage. This is significantly more efficient than the alternative of loading and initializing multiple deep networks for each frame.

Each stage of the framework is implemented as an interface. This allows components in the framework to be removed and replaced with further improved versions of each operation in the framework as the state-of-the-art progresses.

Figure 7: Three consecutive frames showing the original images at the top and the swapped images at the bottom.

Finally, Figure 7 shows three frames resulting from running our framework. The top images show the original frames as captured during shooting, while the bottom shows the same frames using the framework proposed in this work. This prototype of the proposed framework requires minimal user interaction and can produce results suitable for deepfake face replacement for creative projects such as *Virtual Maggie*.

LEGAL CONTEXT

Digital Resurrection vs Deepfakes – a semantic approach or a real division?

A portrayal on screen of a real person is often understood by the audience in the context of the film or broadcast they are watching – an actor playing a role in a biographical film (Gary Oldman as Churchill in *Darkest Hour* (Joe Wright, 2017); Marion Cotillard as Edith Piaf in *La Vie En Rose* (Olivier Dahan, 2007)), an actor portraying a real life person in a clearly fictional film (Janet Baker reprising her impersonation of Margaret Thatcher in *For Your Eyes Only* (John Glen, 1981); Adrien Cayla-Legrand as Charles de Gaulle in *Day Of The Jackal* (Fred Zinneman, 1973)), or real footage of the real life person in a documentary. The audience is normally sophisticated enough to distinguish between these as portrayals and the reality – although during the filming of *Day of The Jackal* Adrien Cayla-Legrand was reputedly mistaken by some members of the public as the real de Gaulle (then dead for two years).

The proliferation of manipulated images and video and audio has excited legal and legislative analysis with a view to potential regulation of deepfakes. Agendas and discussion around the analysis and regulation of such manipulations, without the consent of the original person, have varied from fraud (Metliss and Berggren 2020), to performers rights (Pavis 2020), to image rights, rights of publicity and persona protection (Farish 2020; Perot and Mostert 2020), to the criminal law response to revenge pornography (Crofts and Kirchengast 2019) and privacy rights (Chesney and Citron 2019). Legislation in different jurisdictions has seen regulation of manipulated material in often very narrow areas: non-consensual sharing of intimate images (or so-called revenge porn) in Australia, the creation of manipulated videos designed to influence elections in Texas, and manipulated or retouched images in the fashion and advertising industry

in France and Israel. No attempt as yet has been made to comprehensively regulate such manipulations.

The term Deepfake has no uncontested definition. In the Texas legislation, the definition put forward is: ‘s.1: ‘Deep fake video’ is defined as ‘a video, created with the intent to deceive, that appears to depict a real person performing an action that did not occur in reality’ (SB751, section 1). The legislation is limited in this case to video – not audio or still images – and is not specific in the method of creation, whether by human or AI creation. The offence occurs if one creates or causes such a deep fake video to be made. Chesney and Citron prefer a definition that a Deep Fake is ‘hyper-realistic digital falsification of images, video, and audio’ (2019, 1757). In that definition focus is placed on the outcome and method of creation, whilst in the Texas Election Code, the key element is intention to deceive.

Is the intention to deceive the audience a key component in the creation of the manipulated images/videos? It would seem that many of the examples are done for purpose of parody and creating a meme. Taking the example of Charleston (1995), a defamation case, the falsification of the images to show the faces of actors from the soap opera *Neighbours* on the bodies of porn actors would clearly deceive no reader. However, the Channel 4 *Alternative Christmas Message 2020* showed a hyper-realistic portrayal of the Queen, although it could be argued that when taken with the audio, any critical viewer would have realised that the Queen was not actually giving the address, as was the intention.

There are differing schools of thought as to what an intention to deceive could mean. On the one hand, it is suggested that intention needs to be manifested by a clear calculation that the viewer was to be misled (see e.g. the cases involving Trade Marks such as Re Australian Wine Importers and Mason, Re Horsburgh, and Re Maeder’s Application) – that the viewer of the manipulated image would not realise that it has been manipulated and that this was the intention of the creator. On the other hand, some case law has considered that if deception has taken place, then the test is made out, either because it is self-evident that deception has been intended as a likely consequence of the similarities or that the viewer has from their own perspective been likely to be confused (see in the context of Pharmaceutical goods Potter and Clarke Ltd 1947).

Ekaratne (2020) has set out a helpful taxonomy of types of manipulations which could fall within the deepfakes definition. In every case the image or video has been manipulated without the consent of the subject of the image. The types are:

> Category A: Clearly manipulated images with clearly no subject consent to disseminate: With this type of image, it is clear (by virtue of text or context or both) not only that the image is manipulated but also that the subject did not consent to its dissemination.
> Category B: Clearly manipulated images with unclear subject consent to disseminate: Such an image is clearly a manipulated image, but it is unclear to the reasonable viewer whether or not the subject consented to disseminate it. Unlike with Category A images, viewers may believe that the subject consented to dissemination. Finally, Category C: Ambiguously manipulated images are those that do not have a textual disclaimer of manipulation, and that are also not clearly manipulated owing to context. The viewer may believe these are true (non-manipulated) images depicting the subject in real life. Belief that the subject consented to disseminate is also possible, but the main harms lie in a manipulated image being shown as a true image. (359)

Ekaratne distinguishes between categories on the basis of viewer awareness. If the viewer is aware that an image is manipulated, then there is no intention to deceive, neither because the creator is intending to mislead nor because the viewer is likely to be misled. It is only in Category C that there is an outcome that the viewer is likely to be misled. The film producer wishing to use manipulated images such as in *Virtual Maggie* would need to navigate the choppy waters of category 3. The audience is being asked to accept that the scenes on screen depicting Margaret Thatcher are believable, just as with the Channel 4 *Alternative Christmas Message*, within the context of the reception of screen fiction. The audience is invited to accept that the manipulated image is in fact a true image. The audience is asked to distinguish between the portrayal of a real-life character by an actor with a machine-created performance created by technology using real images to create a wholly manipulated ‘true image’.

A film producer will argue that unlike many instances of ‘fake news’, the digital manipulation is purely for entertainment and artistic purposes, that the viewer does not suffer harm as the producer only intends to mislead for entertainment. But in which case, why not use an actor to depict the real-life person? The whole intention of the film producer is to serve up a

photorealistic version that does lead the audience to be likely to be misled. They are looking to present a real-life person in a fictional context (or indeed a context that is masquerading as a factual context) that can appear to the viewer to be reality. Winick distinguished between manipulated images in which 'no reasonable person would believe that factual information is being conveyed' (1997, 191) and images that had realistic portrayals. It is suggested that a film producer would be striving for manipulations in which a reasonable person would consider that factual information is being conveyed. The film producer is relying on the implied trust and consent of the viewer of the film, who is prepared to suspend disbelief.

Virtual Maggie, in its use of the deceased character of Margaret Thatcher, is in a different context to the manipulations of people still living, who can either consent or potentially take action potentially. Our case also differs from the resurrection of fictional characters played formerly by deceased actors, such as in *Star Wars Rogue One* (Gareth Edwards 2016) where Peter Cushing was recreated digitally for the role of Grand Moff Tarkin utilising footage from *Star Wars A New Hope* (George Lucas 1977), where the studio owns rights in the performance embodied in the previous film. It is more akin to the presentation of certain deceased pop stars 'as live' on stage by use of hologram (such as Prince, Tupac, Elvis Presley, Buddy Holly) where digitisation of their previously filmed performances is presented interactively with a live band. However, in that case the audience is not deceived that the portrayal is actually happening in the now. Where digital resurrection is referred to, it can be seen that it is in the Ekaratne Category C use of digitally created film performances of deceased people portrayed as themselves.

Legal approach to death and reputation after death

Death comes to everyone. With death, however, some legal rights come to an end, whilst others are created and still others are continued as if death had not intervened. On death, the deceased may no longer vote in an election, for example, but may still stand as a candidate in an election if already on the ballot. The Courts will uphold the wishes of the deceased made clear in valid testamentary dispositions. The Government may make laws relating to the deceased, such as the UK's Organ Donation (Deemed Consent) Act 2019, which now requires an opt-out rather than an opt-in for organ donation. A whole set of laws around bodily integrity of the deceased arise on death: autopsies, burials, cremations as well as organ donation. The property of the deceased remains in their ownership and control through their agents

(executors or administrators) until disposed of via their testamentary wishes. Certain types of artefacts created through intellectual property remain in existence until a period of time post-mortem (in the UK, 70 years after death for certain types of Copyright), whilst neighbouring rights such as moral rights also remain personal to the deceased post-mortem. Consent is required from those authorised to act on behalf of the deceased to licence the use of these creative artefacts. However, in the UK there are no rights of publicity in the image of a person similar to those that exist in many States of the USA. Some of these exist post-mortem, such as in Tennessee where the right exists in perpetuity. Rights to sue for defamation also do not survive death on the basis that reputation is personal only and the deceased cannot feel harm as a result of a defamatory statement (Hatchard 1887).

Dissemination of manipulated images without the consent of the subject of the image can result in harm to the subject on an emotional level, and on a reputational level, which may result in financial loss. If the subject were alive, they would be able to sue for financial losses for reputational harm, such in Irvine (2003) for passing off as a result of manipulated image (the loss of a fee) or in defamation as in Charleston, although the Court found there was no defamation in that case. It seems otiose that reputational harm post-mortem is not capable of similar protection. Certainly, financially many deceased persons in the entertainment industry attain significant earnings post-mortem: Forbes magazine publishes a list of top earning dead celebrities annually. In 2019, Michael Jackson grossed $60 million while Elvis Presley grossed $39 million – significant amounts which can be affected by reputational damage (Forbes 2019). If manipulated images are considered sui generis, there is no reason why the law should not afford specific protections to the deceased.

Defamation – time for a rethink?

Film producers seeking to digitally resurrect the dead should tread carefully. Manipulated images which ‘blacken the reputation’ of the deceased may not currently be subject to defamation action in many jurisdictions including the UK, but technological development might and perhaps should prompt a rethink.

The ruling in Hatchard clearly was a product of a different age. Dissemination of manipulated images is now much easier than of a statement of a defamatory nature in 1887. In Hatchard, the statement was an assertion that a trademark was being falsely used. It would have been difficult

to see how this statement could have been widely disseminated beyond advertising in a newspaper. The concept of and protection of reputation was completely different in 1887; now, it is well-established that reputation per se is protected under the Human Rights Act 1998. Indeed, the European Court of Human Rights has accepted that damage to the reputation of a deceased can have reputational impact on family members of the deceased, who can give grounds for a claim in defamation (Putistin 2013). Judge Lemmens summarised:

> This judgment is important in that it accepts that under certain conditions the damage to the reputation of a deceased person can affect the private life of that person's surviving family members. The judgment makes very clear, however, that such a situation will occur only in relatively exceptional circumstances. (Putistin 2013, Paragraph 1)

It is submitted that the damage to financial earnings may be greater and more protected. In the UK, the revision of the law was considered in the 1948 Defamation Committee, where it was proposed that the law should not be changed, whilst in the Faulks Committee report of 1975, there was a recommendation that, for a period of five years after death, specified survivors should be entitled to bring an action limited to a declaration that the matter complained of was untrue and to an injunction, but not for damages. This was not enacted. The topic was raised again during the debates leading to the Defamation Act 2013 but not fully debated. The issue has also been raised in consultations before the Scottish Assembly (2011), the Northern Irish Assembly (2014) and also before the Republic of Ireland Dail (2003). However, in certain jurisdictions post-mortem rights to protect reputation exist: in the Philippines, in the State of Tasmania in Australia, and in the states of Georgia, Nevada and Idaho, where specific protection provides for family members to sue for defamation in respect of the publication of false matters (including images) which tend to blacken the name of the deceased.

With digital resurrection, some film producers create manipulated images which are intended to and have the potential to cause the viewer to have a misleading view that the deceased actually acted in the way depicted and made the statements depicted. By its very nature this ambiguity is designed to suspend the disbelief of the audience. Where this affects the reputation of someone recently deceased and so affects their family members directly, it is submitted that the courts should rethink whether they should not be able to sue for defamation. In the circumstances, ethically (if not yet legally) a film producer engaging in digital resurrection such

as with *Virtual Maggie* should certainly seek the consent of those who might be directly affected.

DISCUSSION

This paper investigates the area of historical deepfake imagery from three perspectives: Creative Screen Practice, Technological Practice and Legal. While seemingly disparate areas of research, this work has highlighted challenges both within the respective fields, and between them. These tensions are examined within the concrete setting of the practice-based project *Virtual Maggie.* Creative screen practice has highlighted quality and filmmaking challenges which are both specific to the historical deepfakes in the *Virtual Maggie* project, but also are faced during the wider process of using deepfakes in film production. This leads to the technical challenges where conventional deepfake methods are unlikely to be suitable due to the limited amount of low-quality imagery available. Underpinning both areas is the legal aspect which examined the reputational and defamation implications of using a historical figure in a new context. The legal aspects discussed in this paper are likely to both constrain and guide filmmaking and technological approaches.

Deepfake systems typically store a representation of the face to be replaced encoded in the parameters of a neural network which can then be used to reconstruct the face in a new pose. The proposed framework in this paper takes a different route, which requires an original image of the actor's face. Both these approaches need to consider the input data; conventional approaches need to consider usage rights and future legal consequences for a large amount of imagery required to train the decoder, whereas the proposed framework requires rights for just a single image.

This leads to creative practice issues as to how well future technology will be able to represent and warp an input historical face into poses which diverge from the captured pose, and the constraints this will impose on filmmakers when planning shoots where actors’ faces will be replaced. As the technology behind deepfakes is progressing very quickly, establishing guidelines for filmmakers is also challenging due to the continuously changing requirements imposed by the technological state-of-the-art.

It should be noted that this paper has focused exclusively on the deepfake image. Machine learning can also be used to create deepfake audio – in comparison a more straightforward process. However, in screen drama the filmmaker's desire to direct an actor's vocal performance would appear to exclude the use of artificially delivered lines of script. No matter how perfectly convincing a deepfake voice might be, many directors would prefer the creative opportunities of collaborating with a voice actor to deliver this key element of a synthetised screen performance.

In summary, this work has found that the three areas examined in this work are deeply intertwined. Legal issues may impact the development and deployment of the technology behind deepfakes, whereas the technology has a significant impact on the constraints and opportunities for filmmakers, who in turn may create content which has consequences for the perceived legacy and reputation of historical figures.

CONCLUSION

Our research during the *Virtual Maggie* project highlights considerable obstacles to the adoption of a deepfakes approach to digital face replacement in High Definition television drama and independent film. This article has highlighted creative and technological issues facing creative practitioners in working with machine learning and has explored the legal issues associated with the digital resurrection of the deceased. The ambition of our study was to investigate whether a deepfakes approach to digital face replacement is viable in high definition screen production. At this point, we can conclude that neither *Virtual Maggie* nor Channel 4's *Alternative Christmas Message* has been able to demonstrate an equivalent believability in digital face replacement to the work of the DVFX postproduction houses for Disney and other major studios. However, the fact that Framestore, a company responsible for the DVFX on big-budget features such as *Mulan* (Niki Caro, 2020) and *Gravity* (Alfonso Cuaron, 2013), is already involved in deepfakes for television broadcast indicates that corporate leaders in this part of the screen industry are convinced of the future for machine learning. Although the task of generating deepfakes of historical figures comes with its own set of technological challenges, our experimental framework developed in the *Virtual Maggie* prototype offers a solution towards solving these limitations though a novel combination of deep learning techniques.

The most important questions still to be addressed are the ethical issues that must be confronted before the point, in a few years' time, when advanced machine learning allows digital

face replacement to be widely accessible at a very high level of believability. In creating *Virtual Maggie*, we have become acutely aware of the responsibilities of creators when working with this technology. With deepfakes entering the mainstream of screen production, there is an opportunity for both legislators and industry stakeholders to address the complex ethical issues that arise from this significant change in screen culture.

References

Alternative Christmas Message (2020) Channel 4 TV. Framestore

Antic J (2020) https://github.com/jantic/DeOldify

Anwar S, Khan S and Barnes N (2020) A Deep Journey into Super-resolution: A Survey. *ACM Computing Surveys (CSUR)* 53: 1–34.

Blanz V, Scherbaum K, Vetter T and Seidel H-P (2004) Exchanging faces in images. *Computer Graphics Forum* 23: 669–676.

Bode L (2007) 'Grave Robbing' or 'Career Comeback'? On the Digital Resurrection of Dead Screen Stars. In: Kallioniemi K, Kärki K, Mäkelä J, and Salmi H (eds) *History of Stardom Reconsidered*. Turku, Finland: Inter-national Institute for Popular Culture

Bode L (2010) No Longer Themselves: framing digitally enabled posthumous performance. *Cinema Journal* 49 (4): 46-70

Bode L (2017) *Making Believe: Screen Performance and Special Effects in Popular Cinema.* New Brunswick: Rutgers University Press.

Cheng Z, Yang Q and Sheng B (2015) Deep colorization. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 7-13 December 2015, pp. 415–423.

Chesney R and Citron D (2019) 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security *California Law Review 1753, U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21.*

Crofts T and Kirchengast T (2019) A ladder approach to criminalising revenge pornography. *Journal of Criminal Law J. Crim. L. 2019* 83(1): 87-103.

Deepfakes (2020) https://github.com/deepfakes/faceswap

Dong C, Loy CC, He K. and Tang X (2014) Learning a deep convolutional network for image super-resolution. *European conference on computer vision*, Columbus, Ohio, 24-27 June 2014, pp. 184–199.

Dong C, Loy CC, He K. and Tang X (2015) Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence* 38: 295–307.

Ekaratne SC (2020) Manipulated images: a taxonomy. *European Intellectual Property Review* 42(6): 353-363

Farish K (2020) Do deep fakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deep fake. *Journal of Intellectual Property Law and Practice: J.I.P.L.P.* 15(1): 40-48

Forbes Magazine (2019) The Top-Earning Dead Celebrities Of 2019: Available at https://www.forbes.com/sites/zackomalleygreenburg/2019/10/30/the-top-earning-dead-celebrities-of--2019/ (accessed 03 January 2021).

Goodfellow I et al. (2014) Generative adversarial nets. *Advances in neural information processing systems:* 2672–2680.

Helm D and Kampel M (2020) Overscan Detection in Digitized Analog Films by Precise Sprocket Hole Segmentation. *International Symposium on Visual Computing,* San Diego, CA, 5-7 October 2020, pp. 148–159.

Horn BK and Schunck BG (1981) Determining optical flow. *Techniques and Applications of Image Understanding* 281: 319–331.

Iizuka S and Simo-Serra E (2019) DeepRemaster: temporal source-reference attention networks for comprehensive video enhancement. *ACM Transactions on Graphics (TOG),* 38: 1–13.

Kazemi V and Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE conference on computer vision and pattern recognition* Columbus, Ohio, 23-28 June 2014, pp.1867–1874.

King DE (2009) Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10: 1755-1758.

Lathuilière S, Tulyakov S, Ricci E and Sebe N (2020) Motion-supervised Co-Part Segmentation. *arXiv* preprint arXiv:2004.03234.

Liu M-Y, Breuel T and Kautz J (2017) Unsupervised image-to-image translation networks. *Advances in neural information processing systems,* 30: 700–708.

Metliss E and Berggren S (2020) Can you believe your eyes? Deepfakes and fraud. *Fraud Intelligence F.I. 2019/20* Dec/Jan: 12-14

Naremore J (1986) Expressive Coherence and the 'Acted Image'. *Studies in the Literary Imagination*, *19*(1), 39-54

Naruniec J, Helminger L, Schroers C et al (2020) High-Resolution Neural Face Swapping for Visual Effects. *Eurographics Symposium on Rendering* 39 (4)

Nazeri K, Ng E and Ebrahimi M (2018) Image colorization using generative adversarial networks. *International conference on articulated motion and deformable objects*, pp. 85–94.

Nguyen TT, Nguyen CM, Nguyen DT and Nahavandi S (2019) Deep learning for deepfakes creation and detection. *arXiv preprint arXiv:1909.11573*

Nirkin Y, Keller Y and Hassner T (2019) FSGAN: Subject agnostic face swapping and reenactment. *Proceedings of the IEEE international conference on computer vision* Long Beach, CA, 15-20 June 2019, pp. 7184–7193.

Pavis M (2020) 'Submission to the UK IPO: Artificial Intelligence and Performers' rights'. Report, The Centre for Science, Culture and the Law at the University of Exeter.

Perot E and Mostert F (2020) Fake it till you make it: an exploration of the US and English approaches to persona protection as applied to deep fakes on social media. *Journal of Intellectual Property Law and Practice: J.I.P.L.P. 2020* 15(1): 32-39

Thies J et al. (2015) Real-time expression transfer for facial reenactment. *ACM Transactions on Graphics (TOG)* 34 (183): 1–14

Viola P and Jones M (2001) Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition.*Kauai, Hl, 8-14 December 2001

Vincent J (2018) 'Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news' https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed

Wilson P I and Fernandez J (2006) Facial feature detection using Haar classifiers. *Journal of Computing Sciences in Colleges* 21: 127–133.

Winick R (1997) Intellectual Property, Defamation and the Digital Alteration of Visual Images. *21 ColumVLA Journal of Law and Arts*: 143

Yuzhi Z et al. (2020) Legacy Photo Editing with Learned Noise Prior. *arXiv preprint* arXiv:2011.11309.

Zhang R, Isola P and Efros AA (2016) Colorful image colorization. *European conference on computer vision 2016.*

Further Legal References:

Statutes

Administration of Estates Act 1925
Copyright Designs and Patents Act 1988
Defamation Act 2005, Tasmania Consolidated Acts
Defamation Act 2013
Georgia Code § 16-11-40
Idaho Code § 18-4801
Nevada Revised Statutes § 200.510
Organ Donation (Deemed Consent) Act 2019
Revised Penal Code of the Philippines, Article 353
SB 751, 86th Legislature Regular Session (Texas 2019) amending Section 255.004, Election Code

Cases

Re Australian Wine Importers and Mason 41 ChD 278
Charleston v News Group Newspapers Ltd [1995] 2 A.C. 65
Hatchard v Mège (1887) 18 QBD 771

Re Horsburgh 53 LJ Ch 237

Irvine v TalkSport Ltd [2003] EWCA Civ 423; [2003] 1 W.L.R. 1576.

Re Maeder's Application [1916] 1 Ch 304

Potter and Clarke Ltd v The Pharmaceutical Society of Great Britain [1947] 1 Ch 483

Putistin v Ukraine 21 November 2013 Application No 16882/03

Reports

Legal Advisory Group on Defamation (2003), Report of the Legal Advisory Group on Defamation, Dublin

Report of the Committee on the Law of Defamation (Cmd 7536) (1948)

The Report of the Faulks Committee on Defamation (Cmd 5909) (1975)

Scottish Government (2011), Death of a Good Name - Defamation and the Deceased: A Consultation Paper.

[i] https://www.youtube.com/watch?v=IvY-Abd2FfM

[ii] https://www.channel4.com/press/news/deepfake-queen-deliver-channel-4s-alternative-christmas-message See also viewer comments on the channel's Youtube (https://www.youtube.com/watch?v=IvY-Abd2FfM)

[iii] Xx, xx, xx (2020) *Virtual Maggie: technological opportunities and ethical dilemmas in the development of virtual performers for feature films and television* (UWE Bristol)

[iv] Comment in *The Hobbit: An Unexpected Journey Extended Edition* 2012 Bluray Warner Brothers

[v] There are several related concepts, such as novel face synthesis, expression swapping, attribute manipulation, all of which are subtly different from swapping entire faces.