

CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Communication Studies and Journalism

Master thesis

2022

Nathalie Weatherald

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Communication Studies and Journalism

Nathalie Weatherald

**Reinventing the ‘blurry oval’?
Practitioner perceptions of deepfakes as a tool
for anonymisation in documentary film and
video journalism**

Master Thesis

Prague 2022

Author: Nathalie Weatherald, BA

Supervisor: Mgr. Sandra Lábová

Academic Year: 2021/2022

Bibliographic note

WEATHERALD, N. (2022). *Reinventing the blurry oval: Practitioner perceptions of deepfakes as a tool for anonymisation in documentary film and video journalism* [Master thesis, Charles University].

Abstract

In 2020, the documentary film *Welcome to Chechnya* disguised the sources it portrayed using deepfake-like 'digital masks', to wide acclaim: many described the use of the technology in this way to be game-changing for the industry. This qualitative study examines documentary filmmakers' and video journalists' (practitioners') perceptions of the benefits and limitations of the use of deepfakes, or AI-assisted synthetic media, to anonymise sources in their work, in the context of theoretical understandings of photographic realism and applied journalistic ethics. Through one unstructured interview with the film's visual effects supervisor, Ryan Laney, and eight semi-structured interviews with practitioners who have previously visually disguised sources, the study identifies four key themes of practitioners' views about the use of deepfakes as a tool to anonymise: the impacts on practitioner-source relations, practical considerations, aesthetic impacts of synthetic media and broader industry implications. Overall, practitioners emphasised the limitations of the potential use of deepfakes in this context, much more than the benefits.

Keywords

anonymity, deepfake, documentary film, synthetic images, video journalism

Abstrakt

Dokumentární film *Welcome to Chechnya* (Vítejte v Čečensku) z roku 2020 maskoval zdroje, které zobrazoval, pomocí "digitálních masek" podobných deepfake. Tato praxe se setkala s velkým ohlasem. Mnozí označili použití této technologie v daném žánru za převratné. Předkládaná kvalitativní studie zkoumá, jak dokumentaristé a videožurnalisté (tvůrci) vnímají přínosy a omezení používání deepfakes neboli syntetických médií s podporou umělé inteligence k anonymizaci zdrojů ve své práci, a to v kontextu teoretického chápání fotografického realismu a aplikované novinářské etiky. Diplomová práce prezentuje jeden rozhovor s vedoucím vizuálních efektů zkoumaného filmu Ryanem Laneyem a osm polostrukturovaných rozhovorů s tvůrci, kteří již dříve vizuálně maskovali zdroje. Na základě takto sesbíraných dat studie identifikuje čtyři klíčová témata názorů tvůrců na používání deepfakes jako nástroje pro anonymizaci. Přesněji dopady na vztahy mezi tvůrci a zdroji, praktické úvahy, estetické dopady syntetických médií a širší důsledky pro průmysl. Celkově odborníci z praxe zdůrazňovali omezení potenciálního využití deepfakes v tomto kontextu mnohem více než výhody.

Klíčová slova

anonymitu, deepfake, dokumentární film, syntetické snímky, video žurnalistika

Range of thesis: 41 pages and 112533 characters

Declaration of Authorship

1. The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.
2. The author hereby declares that all the sources and literature used have been properly cited.
3. The author hereby declares that the thesis has not been used to obtain a different or the same degree.

Prague

30.07.2022



Nathalie Weatherald

Acknowledgments

I am grateful especially to Mgr. Sandra Lábová, for all her support and feedback in earlier drafts; the documentary filmmakers and video journalists who shared their time, experiences and insights for this project, despite difficult and busy schedules; and the staff of the university libraries and cafés of Prague, who facilitated the environments in which this paper was researched and written.

Institute of Communication Studies and Journalism
Approved research proposal

Institute of Communication Studies and Journalism FSV UK	
Research proposal for Erasmus Mundus Journalism Diploma Thesis	
THIS PART TO BE FILLED BY STUDENT:	
Student's surname and given name: Nathalie Weatherald	Registry stamp: / Razítko podatelny: <div style="border: 1px solid black; padding: 5px; margin: 10px auto; width: fit-content;"> <p style="text-align: center;">Univerzita Karlova Fakulta sociálních věd</p> <p>Došlo dne: 15 -11- 2021 -1-</p> <p>Čj: 403 Příloh:</p> <p>Přiděleno:</p> </div>
Start of studies for EMJ (in Aarhus) September 2020	
Your faculty e-mail: 51849004@fsv.cuni.cz	
Study program/form of study: Erasmus Mundus Journalism	
Thesis title in English: Practitioner and Audience Perceptions of Deepfakes in Nonfiction Visual Storytelling	
Expected date of submission (semester, academic year – example: <i>SS 2021/2022</i>) (Thesis must be submitted according to the Academic Calendar.) SS 2021/2022	
Main research question (max. 250 characters): How do practitioners and audiences view the benefits and limitations of using deepfake technology to disguise sources' identities in non-fiction visual storytelling?	
Current state of research on the topic (max. 1800 characters): In 2017, deepfakes—an AI-assisted form of digital manipulation, in which faces can realistically be 'swapped' with other existing or entirely fake ones—began to circulate on the internet. The subsequent four years have seen varied usages of the technology: from videos of politicians delivering speeches they never made (such as Jordan Peele's deepfake of Barack Obama), to the creation of so-called designer porn (the emotional consequences of which were the subject of Noelle Martin's 2018 TED talk), to the animation of photographs of deceased relatives (such as ancestry website MyHeritage's 2021 campaign), to popular face-swap apps used for entertainment on social media (such as ZAO and Reface). Research discourse on deepfakes' relationship to journalism has so far been dominated by issues of trustworthiness and reliability of third-party media, and the corresponding threats to news reporting and distribution (Wahl-Jorgensen & Carlson, 2021; Vatreš, 2021; Westerlund, 2019). However, as noted in some of the more recent literature, a 2020 documentary—David France's <i>Welcome to Chechnya</i> —has demonstrated a novel way in which deepfakes could be a help, not a hindrance, to nonfiction visual storytelling. The film swaps volunteers' faces onto the faces of the film's highly vulnerable subjects, offering a	

solution to the longstanding problem of showing a source's personal testimony with all the emotional nuance of the original, whilst fully disguising their identity. In contrast to other popular usages, the technique was built on informed consent from all participants: the subject, the individuals who had their faces swapped in, and the audience (the film begins with a disclosure statement). This specific usage of deepfakes is yet to be explored in academic literature.

Expected theoretical framework (max. 1800 characters):

The theoretical framework will be built from a combination of prior research on the ethics of deepfakes, as well as literature on the role of nonfiction visual media practitioners and audiences.

The ethical complexities of deepfakes have been approached differently from different disciplines. Political scientists have analysed possible effects on democratic systems (Diakopoulos et al., 2021); identity researchers have discussed the impacts of abuse on women, children and marginalised communities (Yadlin-Segal & Oppenheim, 2021; Ratner, 2021; Ayers, 2021); legal scholars have debated the consequences of regulation or criminalisation (Defino, 2019; Gieseke, 2020).

This research will build on philosophical interrogations of benefits and harms of deepfake technologies, to create a framework which nuances the morality of the technology beyond merely 'good' or 'bad', and instead views their usages as socially, culturally, historically and institutionally situated phenomena whose ethical implications depend on the circumstances of their usage (Silbey & Hartzog, 2019; Bode et al. 2021; Paris & Donovan, 2019; de Ruiter, 2021, Kerner & Risse, 2021).

To frame the anonymity provided by deepfakes in the context of nonfiction visual storytelling, the research will draw from literature on the ethics of visual sources and informed consent, conducted in the context of the social sciences (Wiles et al., 2012; Awad, 2006), as well as literature that addresses anonymisation directly in a journalistic context (Schuhbert, 2015; Scott & Rains 2020). To inform analysis on how audiences respond to deepfakes on screen, the research will examine prior work on the 'digital face' in visual media (Sheldon 2018; Bode 2021; Lees et al. 2021).

Expected methodology, and methods for data gathering and analysis (max. 1800 characters):

The research will collect qualitative data through two methods: semi-structured interviews with practitioners and focus groups with audiences. The design of these studies will be informed by prior literature on conducting them (Cyr, 2019; Galletta & Cross, 2013).

'Practitioners' will be defined as people who work in non-fiction visual storytelling (such as documentary filmmakers, video journalists, photojournalists), who are currently working or have previously worked with sources who must keep their faces disguised.

'Audiences' will be defined as people with an interest in nonfiction visual storytelling (documentary film, video journalism, photojournalism).

Potential interviewees and focus group participants will be identified through purposeful sampling, and contacted via industry forums, directly via email, or through

recommendation. Interview and focus groups will then be conducted online, recorded and transcribed.

Expected research design (data to be analyzed, for example, the titles of analyzed newspapers and selected time period):

Since the use of AI to digitally disguise subjects in nonfiction visual storytelling is a relatively new practice, with as of yet very little practical usage and research around it, this thesis will use an explorative research design. The aim of the study will be to identify some common perceptions of the technique's benefits and limitations, from the perspectives of both the practitioners and their potential audiences, which can be later explored in more depth and nuance by further study.

Expected thesis structure (chapters and subchapters with brief description of their content):

- Introduction
- Literature Review
- Methodology
- Findings
- Discussion
- Conclusion

Basic literature list (at least 5 most important works related to the topic and the method(s) of analysis; all works should be briefly characterized on 2-5 lines):

Bode, Lisa, Dominic Lees, and Dan Golding. 'The Digital Face and Deepfakes on Screen'. *Convergence: The International Journal of Research into New Media Technologies* 27, no. 4 (August 2021): 849–54. <https://doi.org/10.1177/13548565211034044>.

- Issue 4 Volume 27 of the journal *Convergence* is a special issue designed to contribute to the growing body of scholarship on deepfakes and digital faces; this editorial, and the articles in the rest of the issue, will be used to build up a deeper picture of the current discourse around deepfakes and AI on screen.

Paris, Britt, and Joan Donovan. 'Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence'. *Data & Society*, 18 September 2019. https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf.

- This report, published by Data & Society, an independent nonprofit producing original empirical research about emerging technology, argues that deepfakes are no new threat to democracy, and do not change how evidence works: what they do provide is new opportunities for the negotiation of expertise, and therefore power.

Scott, Craig R., and Stephen A. Rains. '(Dis)Connections in Anonymous Communication Theory: Exploring Conceptualizations of Anonymity in Communication Research'. *Annals of the International Communication Association* 44, no. 4 (1 October 2020): 385–400. <https://doi.org/10.1080/23808985.2020.1843367>.

- A synthesis of research on the concept of anonymity across communication studies, across the subfields of journalism, health communication and organisational communication.

Sanders, Willemien. 'Documentary Filmmaking and Ethics: Concepts, Responsibilities, and the Need for Empirical Research'. *Mass Communication and Society* 13, no. 5 (29 October 2010): 528–53. <https://doi.org/10.1080/15205431003703319>.

- A review of the discourse on documentary filmmaking and ethics, particularly regarding the filmmaker-filmed relationship.

Wahl-Jorgensen, Karin, and Matt Carlson. 'Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes'. *Journalism Practice* 15, no. 6 (3 July 2021): 803–20. <https://doi.org/10.1080/17512786.2021.1908838>.

- An investigation into journalistic discourses on deepfakes as the future of fake news, over an 18-month period.

Related theses and dissertations (list of B.A., M.A. and Ph.D. theses defended at Charles University or other academic institutions in the last five years):

Poon, Jessica. 'GANs gone wild: Public perceptions of Deepfake technologies on Youtube'. Univerzita Karlova, 2020. <https://dspace.cuni.cz/handle/20.500.11956/150489>

Sheldon, Zachary. 'Digital Characters in Cinema: Phenomenology, Empathy, and Simulation'. Baylor University, 2018.

<https://baylor-ir.tdl.org/bitstream/handle/2104/10362/SHELDON-THESIS-2018.pdf?sequence=1>.

Date / Signature of the student:

[Redacted signature]

15.11.2021

THIS PART TO BE FILLED BY THE ACADEMIC SUPERVISOR:

I confirm that I have consulted this research proposal with the author and that the proposal is related to my field of expertise at the Faculty of Social Sciences.

I agree to be the Thesis supervisor.

15/11/2021	[REDACTED]
Surname and name of the supervisor	Date / Signature of the supervisor
Further recommendations related to the topic, structure and methods for analysis:	
Further recommendations of literature related to the topic:	

The research proposal has to be printed, signed and submitted to the FSV UK registry office (podatelna) in two copies, **by November 15, 2021**, addressed to the Program Coordinator. Accepted research proposals have to be picked up at the Program Coordinator's Office, Mgr. Sandra Štefaniková. The accepted research proposal needs to be included in the hard copy version of the submitted thesis.

RESEARCH PROPOSALS NEED TO BE APPROVED BY THE HEAD OF ERASMUS MUNDUS JOURNALISM PROGRAM.

Contents

INTRODUCTION	13
1. LITERATURE REVIEW	17
1.1 Ontology of the image.....	17
1.1.1 Mechanical cameras and early photography.....	17
1.1.2 From picture-making to data-collection device.....	19
1.1.3 Computational photography and synthetic images.....	21
1.1.4 Documentary filmmaking and ‘the charge of the real’.....	22
1.2 Journalistic ethics in practice.....	23
1.2.1 Applied ethics in journalist-source relations.....	23
1.2.2 Navigating a post-truth environment.....	26
1.3 Hybrid modalities and industry 4.0.....	27
1.3.1 Synthetic media on screen.....	27
1.3.2 ‘Smart storytelling’.....	29
2. METHODOLOGY	31
2.1 Research aims and design.....	31
2.2 Data collection.....	31
2.3 Data analysis.....	35
2.4 Refinement of initial themes.....	36
3. FINDINGS AND DISCUSSION	40
3.1 Findings.....	40
3.1.1 Practitioners’ responsibilities to sources.....	40
3.1.1.1 Collaboration as key in maintenance of trust.....	41
3.1.1.2 Imperative of safeguarding.....	42
3.1.2 Practical considerations.....	44
3.1.2.1 Resource constraints.....	45
3.1.2.2 Editorial oversight.....	45
3.1.2.3 Deepfakes opening up new avenues of storytelling.....	45
3.1.3 Aesthetic impacts of synthetic media.....	46
3.1.3.1 Preference for in-camera anonymisation.....	46
3.1.3.2 Blurring the boundaries of the format.....	47
3.1.3.3 Narrative impact of masks.....	48
3.1.4 Industry implications.....	49
3.1.4.1 Novelty of technology.....	49
3.1.4.2 Rupture trust with audiences.....	50
3.1.4.3 A slippery slope.....	51
3.2 Discussion.....	52
5. CONCLUSION	54
6. BIBLIOGRAPHY	55
7. APPENDIX	60

Introduction

On 26 January 2020, the documentary feature film *Welcome to Chechnya* (France, 2020) premiered at Sundance Film Festival in Utah. The film followed the journey of a group of LGBT refugees and survivors of torture fleeing the republic of Chechnya, a part of the Russian federation, during the anti-gay purges of the late 2010s. The main characters of the film – the refugees – were the centre of the story, the cornerstone of its emotional charge, as they recounted to camera harrowing tales of persecution acted upon them by the regime of Ramzan Kadyrov, the region’s leader. Hidden phones and GoPros were used to film their daring escapes from their country, through a network of safehouses and clandestine travel routes. The refugees were – and remain – at high risk of retribution from the state, and the film was produced on the condition of their absolute anonymity: a gold standard in which “even their own mothers would not recognise them” (WITNESS, 2020, 48:36)

In the lead up to the screening, the director, David France, and visual effects supervisor, Ryan Laney, circulated with clips of the film on their phones, playing them to industry delegates to gather feedback on their approach to anonymising the characters on screen. Laney had been brought onto the project by France, who felt dissatisfied with using legacy approaches to anonymisation: the blurry oval, pixelation, shadows. Working in windowless editing rooms, using offline hard drives and computers, the two had designed an entirely new approach, the idea of which was to create “a digital prosthetic where 100 percent of the motion, the emotion, and the essence of what the subject is doing is there” (Heilweil, 2020). Using deep learning technology, the team created deepfake-like digital masks to transplant volunteer activists’ faces onto those of the original subjects of the film.

Deepfakes are synthetic media, in which deep learning – a form of artificial intelligence (AI) – is used to convincingly ‘swap’ the face or voice of a person in a video or image for that of another. Described as “the 21st century’s answer to Photoshopping” (Sample, 2020), deepfakes were initially used to fake pornographic videos or create false videos of public figures appearing to say and do things they did not say or do. Since 2017, when the first synthetic images of this kind emerged, the bar of the requisite technical know-how and cost has lowered: multiple smartphone apps and websites exist that generate a deepfaked image within a few clicks. Correspondingly, the images have become more realistic, prompting widespread concern on the disruptive potential of the technology in an electoral context (Diakopoulos & Johnson, 2021; Shin & Lee, 2022), challenges for detection in a legal and regulatory context (Cover, 2022; Fernandez, 2022), as a tool for gendered violence (Lucas, 2022; Newton & Stanfill, 2020; Popova, 2020), and deepfakes’ impact for truth and trustworthiness in society as a whole (Adjer et al., 2019; Gosse & Burkell, 2020).

Industry delegates, having been shown the short clips, reacted positively, said Laney (Appendix). The film proceeded to win Sundance's Special Jury Award for Editing, the first of a string of prizes. In critics' circles, the documentary received wide acclaim: the use of deepfake-like technology to anonymise sources was described as a "game changer" by multiple publications (Grobar, 2021; Thomson, 2020). In interviews, France said he preferred to refer to the technology developed by his team as "deeptruths", as the purpose is not to deceive, but to reveal (WITNESS, 2020). Laney set up Teus Media, a new company founded to support journalistic storytelling and provide identity preserving treatments that help protect the privacy of people who cannot be seen on camera (Teus Media, 2022). The company has subsequently consulted on over fifty future projects on topics from China and North Korea to dark money and human trafficking (Appendix).

While a long way from becoming part of the standard repertoire of documentary filmmakers' and video journalists' toolkits, the interest in Teus Media's services and degree of media hype surrounding the film indicates a growing potential of the use of deepfakes in this context. As such, the purpose of this study is to conduct exploratory research into how video journalists and documentary filmmakers (hereafter referred to as 'practitioners') perceive the benefits and limitations of the use of deepfakes, or similar AI-assisted technology for synthetic image creation, as a tool for the anonymisation of sources on screen. Protocols for why, when, how, and to whom, anonymity is granted in a journalistic context is governed by a multiplicity of industry standards, institutional practices, individual journalistic choices and security considerations. The creation of an exhaustive map of all such interplaying factors is not the focus of this research. Rather, through specific examination of practitioners' perceptions of the use of deepfakes in this context, the study seeks to illuminate aspects of the contested relationship between the truth claims of non-fiction filmmaking, synthesised media and applied journalistic ethics, and in doing so, outline potentially fruitful avenues for future research.

The first chapter, the literature review, begins by locating visual media within theoretical understandings of the ontology of the image. The first section maps the development of realist conceptions of the 'truth value' of photographic images, from the view of the camera as a mechanical, light-recording object, to the camera as data-collection device, to innovations in computational photography and synthetic images, combined with how such conceptions echo through documentary practice. The second section explores journalistic ethics. It begins with a look at applied ethics in journalist-source relations framed within the narrative power of the 'talking head' and frameworks for navigating visual anonymity. Then, ethical considerations are explored within the broader contexts in which they operate, through an overview of how journalists navigate 'post-truth', highly digitised environments – and how deepfakes themselves problematise and undermine reporting in such

contexts. The literature review concludes with a brief overview of some prominent uses of synthetic media on tools and formats are being integrated into contemporary creative documentary practice.

The second chapter, the methodology, details the research aims and design, data collection methods and analysis of the study. Qualitative data was gathered through eight semi-structured interviews with practitioners who have previously visually anonymised sources in their work using legacy methods such as blurring or pixelation. The interview design was informed by an unstructured interview with Laney (Appendix) on his experiences of producing deepfakes as digital masks for journalistic purpose. They involved some questions on prior experiences with source anonymisation, followed by the screening of a clip from *Welcome to Chechnya* (France, 2020), to demonstrate the use of deepfakes to anonymise, followed by a discussion on practitioners' initial reactions to the technology in this context, whether it is something they might consider using, and how they would feel if it became more widespread across the industry. The interviews were then analysed thematically, which resulted in the formation four themes: practitioners' responsibilities to sources, practical considerations, aesthetic impacts of synthetic media, and industry implications.

The third and final chapter lays out the findings and analysis: overall, practitioners emphasised the limitations of such technologies, and reasons they would be unlikely to integrate it into their practice, with relatively little reference to the benefits. As such, practitioners demonstrated that while new software capabilities and hybrid modalities are inevitably increasingly converging with journalistic practice, preference for limited manipulation and concern over the 'slippery slope' of the integration of AI-assisted technologies prevail.

The initial idea of the thesis was to explore the use of deepfakes as a tool for visually anonymising participants from the perspective of both practitioners and audiences. The main reason for shifting the focus of the research from both perspectives to purely those of practitioners was one of scope and feasibility. For one, the process of identifying and contacting practitioners turned out to be far more time-consuming than initially expected. It became apparent early on that to simultaneously conduct focus groups to such a standard that they would deliver high-quality and meaningful data would involve some form of compromise, either in the quality of the practitioner interviews, the focus groups, or both. Data-gathering from practitioners was prioritised over that of audiences for two reasons. Firstly, to satisfy the research goal – an exploratory outline of perceived benefits and limitations, the decision was made that the experience that practitioners could speak from would be more data-rich than that of audiences. As there are so few pre-existing examples of deepfakes being used in this context, questions of whether that can be attributed to resource and accessibility factors, or some other form of principled resistance to using synthetic tools in their work, emerged as an interesting factor to examine in relation to benefits and limitations; this was information that

practitioners would be much better placed to speak on than audiences. Moreover, through the qualitative semi-structured interviews with practitioners, more variation was expressed than expected, and differences were observed between those who primarily viewed themselves as journalists, in contrast to those who viewed themselves as primarily as filmmakers. Instead of casting the data-gathering net broader, this study chose to focus on the nuances in the data collected from practitioners, to give space for full and detailed analysis of the themes that emerged, as opposed to a cursory sweep over a larger quantity of data.

Naturally, eliminating audience perspectives and narrowing the scope of the initial research question created limitations. In chapter 3.2, the discussion, the boundaries of what falls outside the scope of this study are explicitly addressed, supported by suggestions for further research avenues that have emerged throughout the process of conducting this study.

1. Literature Review

The first part of this chapter seeks to outline key theoretical developments in ontological understandings of the photographic image. Debate regarding photography's relationship to 'true' representation is as old as the medium itself (Snyder & Allen, 1975); as such, the chapter begins by rooting contemporary discussion on synthetic images in the context of early conceptualisations of photographic realism, as articulated by Andre Bazin, before exploring how the advent and mass-uptake of digital photography complicated such understandings in the collision of old and new practices of image-making. Moving from digital image-making to synthetic images, the chapter explores how AI-assisted visualisations further challenged ideas of 'seeing is believing', before exploring how such ideas of truth-claims are framed and understood within the context of documentary practice.

The second part of this chapter broadens out from abstract understandings of 'truth' to locate them in practice. The first section focuses on journalist-source relations, examining the power of talking heads and first-person testimony in witness media, before locating such relations in ethical frameworks offered by visual ethnography. The second section looks more broadly at challenges facing journalists in navigating contemporary 'post-truth' environments: how journalists posit themselves as 'truth-tellers' and how this manifests through industry understandings of accepted levels of manipulation, how deepfakes conjure conjectures of 'fearful futures' that necessitate regulation and detection in order to protect journalism from loss of credibility.

The third part takes a more detailed look at synthetic media on screen, and the ways in which AI-assisted technologies have been used to augment storytelling processes, with the aim of provoking discussion and providing entertainment. The chapter concludes by contextualising such 'hybrid modalities' within the context of 'Industry 4.0' and looking at how such developments challenge the norms of the form.

1.1 Ontology of the image

1.1 Mechanical cameras and early photography

Theoretical debate on the status of the image as evidence can be broadly cast into two camps. The first is defined by an essentialist, realist understanding of the image – the photograph constituting a mirror held up to the world, a purely mechanical process in which a moment in time is recorded, imbuing with resultant frame with evidentiary – analogue – value. The second favours the image as a

constructed product – the result of a complex interaction of choices, processes, conventions and assumptions, both at the stage of its creation, and also the contexts in which it is printed, exhibited, viewed and understood.

Realist film theorist Andre Bazin's landmark essay, *Ontology of the Photographic Image* (Bazin, 1960), laid out the case for the former. In his text, Bazin separates the purposes of the visual arts into two aims: those driven by aesthetics and those driven by the psychological desire for (re)presentation. The invention of photography in the 19th century, he argues, freed painting from fulfilling the desire for the latter, thereby opening up the possibilities for abstract art – and shifting onto the medium of the photographic image the charge of representation of the real. Bazin's assertion is that this realism is rooted in the *automatic* and *mechanical* nature of photography, that the light itself – and the non-living object of the camera – is the principle author of the image. While Bazin acknowledges the agency of the photographer in the selection, orientation and approach of the image, he argues that the degree of agency a photographer exerts on the creation of the image is not alike to that of a painter and their image. “The image may be out of focus, distorted, devoid of colour and without documentary value; nevertheless, it has been created out of the ontology of the model. It is the model,” Bazin writes (Bazin, 1960, p. 8). Therefore: even ‘bad’ photographs still contain the *being* of the object presented – it “transfers the reality from the object depicted to its reproduction” (p. 5). Moreover, he emphasises the effect on the viewer of the photograph – that no matter the critical faculties of the viewer, photography possesses an “irrational power [in which] we believe without reservation” (p. 8).

Bazin here obliquely invokes pragmatist philosopher and semiotician Charles Sanders Peirce's categorisation of a sign into three phenomenological categories: an icon, symbol or index (Peirce, 1985). A sign's iconic value lies in its physical resemblance to the object – in other words, it has enough fidelity of representation to be understood as the signified object that it does not need further contextual explanation or socially-informed understanding to be broadly understood as referring to that object. A sign's symbolic value, in contrast, necessitates context or societal convention in order to be understood. It has no resemblance between the signifier and the signified – for example: road signs, national flags or letters of the alphabet. A sign's indexical value is one of *evidence*. Its existence is testament to the existence of the object, or event, it signifies, such as a footprint, a bullet hole, or woodsmoke. While not necessarily *resembling* the object, it makes a claim to the concrete existence of that object. As such, photographs tend to combine both iconic and indexical qualities. Inherent in a photograph is a resemblance to the original object – transferred, as Bazin described, through the automatic means of the camera – and a testimony to the, even if momentary – existence of that object, or scene. Bazin's argument is that even if a photograph was to have lower iconic qualities, through

being blurred or distorted, a photograph, in its essence, contains a compelling degree of indexical value.

Photography's indexical value has been the subject of exploration by cultural critics, philosophers, theorists, since its invention. In parallel to realist understandings of the function of photography, emerged the alternative camp: one that foregrounded the photograph as a contextualised object, existing within a myriad of structures, messages, biases and assumptions. Foundational photographic texts have pondered the image's contested relation to its referent (Berger, 1973; W. J. T. Mitchell, 1986; Ritchin, 2009; Tagg, 1993). In Roland Barthes' *Image-Music-Text* (1977), Barthes describes photographs in the press as messages, understood through codes: "The structure of the photograph is not an isolated structure, it is in communication with at least one other structure, namely the text – title, caption or article – accompanying every press photograph" (p. 16). In *Camera Lucida* (1982), Barthes reflects on the camera's testimonial power over time and death, as he processed the grief of his mother's passing. Likewise, Susan Sontag's *On Photography* (1978) reflects on the authority conferred by the image: "A photograph passes for incontrovertible proof that something happened" (p. 3). And yet, she writes: "Despite the presumption of veracity that gives all photographs authority, interest, seductiveness, the work that photographers do is no generic exception to the usually shady commerce between art and truth" (p. 4; for more on the debate between photography as journalism or art, see: (Allen, 1986; Deschin, 1960; Postema & Deuze, 2020).

1.2 From picture-making device to data-collection device

Digital photography – and the softwarisation of image-creation as a whole – further complicated theoretical understandings of the iconic, or more presciently, indexical value of a screen-based image. The 'digital revolution' amplified problems in the categorisation of photography and moving image in its many guises and purposes; evolution of technical possibilities have developed in symbiosis with cultural, political, economic, and social processes of change, "compel[ling] practitioners and scholars to rethink stale analogue-photography based ethics" (Mortensen, 2021, p. 1).

In the introductory essay of the first edition of Martin Lister's *The Photographic Image in Digital Culture* (Lister, 1995), Lister responded to the "epochal change" brought to the photographic world by technological developments and the growth of an image-based economy, reflecting on a culture "pervaded by a heavy mixture of new millenarian futurology, the visionary excesses of postmodern thought, and of utopian promise and cultural pessimism" and how such dynamics acted on the ruptures in visual culture (p. 1-4). The invention and marketing of new devices to capture, store, process, edit and distribute digital images in a newly networked internet context challenged

realist assumptions of photographic truth at every stage. In World Press Photo's (Campbell, 2014) report on image manipulation, Campbell emphasises the erroneous logic that a camera, in the digital age, is still a "picture-making device", describing instead how it functions as a "data-collection device": CCD/CMOS sensors record light intensity in comparison to grayscale, with a colour filter array accordingly constructing the colours recorded through computational processes, recorded as a data-set in a RAW file, and subsequently converted by in-camera or photo-editing software into a JPEG, or observable image (p. 7). As such, Campbell specifies, all digital photography is by necessity "post-processed" – no true, original, image exists (p. 9).

Moreover, the broad accessibility of photo-editing software made alteration of digital imagery the domain of not only photographers, but almost anyone: the release of Photoshop 1.0 in 1990 was closely associated with the corresponding threat to photographic truth (Lister, 2013, p. 13). "Press photographers scented a cybernetic dystopia in the making" wrote William Mitchell, professor of arts and sciences at MIT, "a world infested with subversive, uncontrollable imager hackers who would appropriate photographic fragments at will and recombine them into fictions" (W. J. Mitchell, 1994, p. 16). He concludes, succinctly: "the referent has come unstuck" (p. 31). Some, however, view the challenges and processes of digitalisation in less absolute terms. Regardless, "photographic realism remains a significant standard for digital imaging as both a technological and cultural issue," David Bate writes in his reflection of the impact of the digital condition on photographic image-making (Bate, 2013, p. 86). Bate continues: "Even if the ontological base of analogue and digital forms are different, their epistemological problems are the same: they seem to show something that is incontrovertibly 'there' and thus give a strong continuity in the ideological function of digital photographic pictures inherited from analogue photography" (p. 87).

Lister, in the introductory text of the first edition of *The Photographic Image in Digital Culture* (1995), also took issue with these more purist conceptions of photography, criticising the essentialist assumption that "digital photography simply broke the photograph's indexical connection to its reference and that a digital photograph was not (or could not be) indexical" (Lister, 1995, p. 3). His reasoning was that all photography has always interacted with other technologies – print, graphic, electronic, televisual and telegraphic – and that "analogue photographs were intertextual and polysemous and [...] these were not newly defining or distinguishing qualities of digital images" (p. 3). The introductory to the second edition of the same book, written twenty years later, took a slightly more mollified stance. "We should remember to keep [photography's] plurality or multiplicity of forms and uses in view; we should keep its indexicality within strict critical limits; we should be aware of the enormous weight of the representational conventions that it embodies while insisting on its (historical as well as current) hybridity and promiscuity with other technologies and practices" (Lister, 1995, p. 3).

1.3 Computational photography and synthetic images

If the advent of digital photography weakened the foundations of photographic realism, AI-assisted synthetic images brought down the whole building. Where digital photography decentred the photographer, by making the photographer and computer-camera ‘co-authors’ of an image, algorithmic tools offer the possibility to completely remove the human agent from the process: for example, in CCTV cameras, drones and satellites (Zylinska, 2017). Beyond moving a person’s hand away from the shutter, computerised *generative* capabilities have opened up a whole new vein of synthesised image-creation, for a variety of purposes and ends. In February 2019, a new website emerged, called This Person Does Not Exist (NVIDIA, 2022). Each time the website is loaded, it generates a fully automated and synthesised ‘human face’. The website, created primarily for entertainment purposes, is a provocative tool to showcase contemporary capacities of artificial intelligence – its engagement value resting on the supposition of surprise that something so ‘realistic’ can be generated by a computer. “People tend not to think about the effect that neural networks have on our lives [...] if AI can create faces for itself [...] then what is going to happen next?” the website reads.

Three years earlier, Justus Thies and his research group presented a tool called ‘Face2Face’ at the Conference on Computer Vision and Pattern Recognition in Las Vegas, Nevada. Thies and his colleagues demonstrated through a live setup how the tool could transfer a source actor’s face onto an individual in a target video in a photo-realistic fashion, in real time – one of earliest examples of what is now known as ‘deepfake’ technology (Thies et al., 2016). Deepfakes – a form of machine learning technology that can be used to create realistic video or audio files of individuals doing or saying things they did not necessarily say or do (de Ruiter, 2021, p. 1312) – and the AI face generator of This Person Does Not Exist – are powered by generative adversarial networks (GANs). GANs consist of two competing neural networks. One is generative: once fed large sets of data samples – for example, thousands of images of a politician or public figure available online – and this extensive amount of visual material is processed in a way that enables the algorithm to ‘learn’ patterns and concerts of facial features and ‘create’ a new face, or mask. The generative network then competes against a second network, applied to evaluate the accuracy of the synthesised photo or video against existing unaltered visuals of the same face, creating an inbuilt process of “automated self-criticism” (de Ruiter, 2021, p. 1315).

As such, in the past five years since deepfakes came into circulation, they have become increasingly realistic, challenging the very notion of “seeing is believing” (Shin & Lee, 2022, p. 412). A Pew Research Center survey found that 63% of Americans perceived made-up or altered videos and images to create “a great deal of confusion about the facts of current issues and events” (A. Mitchell

et al., 2019). Paris and Donovan (2019) reflected on the implications of the increasing sophistication of deepfake technologies in a report entitled ‘Deepfakes and Cheap Fakes’, in which they formulated a spectrum – at which advanced experimental machine learning techniques exist on one end, and other forms of audiovisual manipulation, which rely on cheap, accessible software, or no software at all, exist on the other. The authors respond to the growing climate of alarm in journalistic and academic circles over the deepfake phenomenon, arguing that what these fearful conjectures miss is that the ‘truth’ of audiovisual content “has never been stable” (Paris & Donovan, 2019, p. 2). Truth, the authors posit, has always been socially, politically and culturally determined – and is able to be manipulated by deepfakes and cheap fakes alike. Their assertion is qualified, however, with two important caveats: firstly, the image-saturated and social media-led culture of the present day means that its fairly easy to create a false, synthesised likeness of almost anyone with a public social media account; secondly, the speed and scale of today’s online platforms enable audiovisual manipulations to be transmitted to a wide network of viewers – and thereby make impressions on a larger number of people – than ever before.

1.4 Documentary filmmaking and ‘the charge of the real’

The Oxford Reference definition of documentary is a piece of media “dealing with factual material rather than fictional material, usually with a goal to create new insight or exposure to facts” (Doyle, 2011). Built into the fabric of documentaries, a non-fiction moving images, is some degree of a truth-claim, however contested (see: Winston, 2000). “The contradictions are rich: on the one hand the postmodern deluge of images seems to suggest that there can be no a priori truth of the referent to which the image refers; on the other hand, in this same deluge, it is still the moving image that has the power to move audiences to a new appreciation of previously unknown truth,” writes film Linda Williams (1993, p. 10) Gershon and Malitsky (2010) argue that the rise of post-structuralism and postmodernism contributed to the “unsettling” of the truth-claim of documentary filmmaking. They describe the difficulty of documentarians to “retain the political purchase of claiming the real while acknowledging the postmodern recognition that truth is socially constructed, in part through filmic representation” (p. 65).

In her book *Carnal Thoughts: Embodiment and Moving Image Culture*, American cultural critic Vivian Sobchack draws on film and media studies, cultural studies, and existential philosophy to articulate a “materialist – rather than idealist – understanding of aesthetics and ethics.” (Sobchack, 2004, p. 3). Published in 2004, well before the invention of AI, the contours of debate she outlines around the blurring of the fiction/non-fiction boundary in cinema are not so dissimilar from those re-hashed around the impacts of the integration of synthetic technologies and tools into creative practice. In her writings, she argues for a distinction between the *not real* and the *irreal*. Where the *not real*

stands in contrast to cultural and historical senses of what is real – rendering it impossible, fantastical or implausible, in our relations with the *irreal* “we do not posit real existence so as to then make a judgement about the reality of what we see; instead, the real is ‘bracketed’ and put off to the side as a noncriterion of the work’s meaning, coherence or plausibility” (p. 258).

Building on Bazin’s ontological understandings of the indexical value imbued by the camera, Sobchack notes that “fiction and documentary, as supposedly different logical types as genres, are both reducible to the same logical type as cinematic images,” (p. 259) asking to what extent an audience can tell the difference between them. Sobchack, speaking from a perspective of existential phenomenology, argues that this delineation rests on “an experienced difference in our mode of consciousness, our attention toward and our valuation of the cinematic objects we engage” (p. 261). In other words, truth-value is less ascribed by the textual features of a film or cinematic object, but how those who watch the film experience and appraise it – which, in itself, is a process co-created by the multiplicity of other film-images, genre tropes and culturally-informed perspectives that have been absorbed by a viewer, forming social and cultural frameworks that shape how the non-fiction moving image is constructed and received. Sobchack recites how the blending of the real and the irreal – not to distort, or “lie” to the viewer, but to augment the storytelling process and the audience’s experience – has been a part of cinematic culture for decades “to advance and complicate the verisimilitude of their narratives as well as the viewer’s hermeneutic enjoyment” (p. 263).

“The charge of the real,” Sobchack writes, “is also, if to a varying degree, an *ethical charge*: one that calls forth not only response but also responsibility – not only aesthetic valuation but also ethical judgement” (284). What Sobchack addresses, and seeks to carve a space of analysis for, is an understanding of the complicated space between pure fiction and pure non-fiction and the blurring and contradictions of that boundary: the documentary elements and styles that dramas and comedies employ, such as the use of real archive clips, blended into a fictional feature, or the whole genre of mockumentaries, and the fictional narrative devices employed to augment the visual product of a documentary, such as re-enactment or docudrama.

1.2 Journalistic ethics in practice

1.2.1 Applied ethics in journalist-source relations

The designs, formats and production approaches in which a source’s audiovisual testimony is made compelling and imbued with authenticity – choices that, paradoxically, create distance from the ‘raw’ version of the testimony – have been examined through multiple disciplines. Witness

scholarship, a field informed by cultural and political memory studies, is a domain that can provide texture and insight into the ways in which sources, subjects or witnesses are imbued with, or stripped of, credibility. Researcher and historian of European post-socialism, Sara Jones, examines the linkage between the power of witnessing to concepts of authenticity, and the ways in which such concepts are mediated through audiovisual content. “Memory must be mediated if it is to be transmitted across time and space,” Jones (2017) states, continuing: “[B]ut what is the impact of this mediation and (how) does it create new forms of authenticity?” (p. 136). Considering how the positionality of an eyewitness to a significant event still holds a hegemonic place in public representations of the past (Sabrow, 2012, p. 22, cited in Jones, 2017, p. 135), Jones’ article questions how the mediation effects of the recording, storing, exhibition and distribution of witness testimony acts on the perceived authenticity of the subject. Frosh and Pinchevski (Frosh & Pinchevski, 2011) have described ‘media witnessing’ as the “witnessing performed *in, by, and through* the media [...] the systematic and ongoing reporting of the experiences and realities of distant others to *mass audiences*” (emphasis in original) (p. 1). Jones finds that first-person testimony – or the ‘talking head’ format, in which individual eyewitnesses tell their personal stories in relation to the specific topic at hand, has become common-place, “gives a face” to the otherwise abstract narrative, and “offer[s] a figure of identification for the viewer” (p. 144). Indeed, Jones describes how the phenomenon has become “canonical” in the genre to the extent that actors have been used where ‘real life’ witnesses cannot be found (Classen, 2012, cited in Jones, 2017).

If the value of visual media lies in the richness and texture of “the enormous amount of information [...] they encode in a single representation [...] which ultimately accounts for its power to engage us.” (Grady, 2004, p. 20), then accordingly, such rich information must be responsibly handled. Scholarship on conducting visual ethnography provides a useful insight into the tension between the desire to publish unaltered visual material for its information-value and capacity to make and substantiate argument and the imperative of upholding ethical principles regarding anonymisation. Pink (2013) emphasises that as ethical guidelines and approaches are highly-context based, localised and culturally contingent, meaning “general methods of preventing harm to participants may not be locally applicable,” and ‘good practice’ requires the “the personal, cultural and ethical sensitivity of the ethnographer” (p. 63-64). Central to developments in visual ethnographic scholarship is a shift from uni-directional to collaborative research, in part in an attempt to move away from exploitative forms of knowledge-extraction (Hammersley & Atkinson, 2019). The use of visual media, whilst birthing a whole host of extra ethical considerations as opposed to say, the written text, has been viewed by some researchers as an asset, as it provides a solution to the notion of “giving something back” (Lassiter, 2005). “Field notes and papers are of little use or interest to most participants [...] however, videos and photographs are usually of interest to the people featured in them,” Pink writes (p. 65), adding that visual work can be something in which both subjects and

researchers invest. Central to the practice of collaborative research, however, is more than a simple tit-for-tat exchange of goods – this, as delineated by anthropology scholar Glenn Hinson, constitutes “reciprocal ethnography” (Lassiter, 2005, p. 17). *Collaborative* ethnography, in contrast, entails a constant ongoing discussion, in which a project is “co-conceived” by both the researcher and the researched, with “constant mutual engagement at every step of the process” (p. 17).

However, guidance for social science researchers on when and how to anonymise visual depictions of sources remains “rather general and limited”, according to Wiles et al. (2012, p. 42). Through focus groups and qualitative interviews, the authors describe how research participants “identified the issue of anonymity as a key ethical challenge for visual research” (p. 44). On protecting vulnerable participants, one childhood researcher recounted in the focus group: “you have the real dilemma of how you get across your data [...] it’s about where your conclusions have come from and if a lot of that has come from visual data and we can’t display that it gets very difficult. So, we’ve gone along a very expensive option of commissioning further interpretations of the data that makes them anonymous... so we’re making films and animating that and using actors” (Wiles et al 2012, p. 47). Respondents in the study also highlighted the responsibility of researchers to anticipate the often-unanticipated implications of the research, once published and distributed, and that later problems may arise even with full and considered consent at the time of participating in the research (p. 47).

The question of anonymity – when, how, for who, with what impact – has always been relevant to journalistic practice. Scholarly attention has been paid to the many ways in which the concept reverberates around the profession: in its business value, in relation to journalistic bylines (Arrese, 2021); the impact of online comment sections (Santana, 2014; Wallsten & Tarsi, 2016); how the granting of anonymity factors into negotiations between journalists and interviewees (Bakker et al., 2013); and the implications on power dynamics in journalist-source relations (Peters & Broersma, 2013). Research on anonymity in journalism and media is a fragmented field: in Scott & Rains’ (2020) exploration of conceptualisations of anonymity in communication research, the authors find that “research on anonymity tends to be isolated within the silos that mark individual subfields, and anonymity is often studied as one small factor among many other more important variables” (p. 1). In Kimball’s (2011) study of the processes by which journalists grant anonymity to sources, she finds that “journalists said protecting a source’s identity is one of the ethical principles to be honoured most strongly” (p. 37), though is qualified by a number of considerations, including intense vetting of the credibility of sources, conversations with supervisors, and legal cover. The Associated Press’s guidelines for the granting of anonymity stipulate three conditions: that the material is information and not opinion or speculation, and is vital to the report; that the information is not available except under the conditions of anonymity imposed by the source; that the source is reliable, and in a position

to have direct knowledge of the information (Associated Press, n.d.). Scott and Rains advocate for further research on anonymity in journalism, broadened from the narrower conceptual approaches of the levels to which anonymity is deployed (Pjesivac & Rui, 2014) to why it is employed (Duffy & Freeman, 2011; Gladney et al., 2013), to “the ways in which digital journalism contributes to reader perceptions of source confidentiality [...] and consider the dimension of visual anonymity.” (p. 10).

1.2.2 Navigating a post-truth environment

One avenue for interpreting the translation of understandings of which forms of visual content constitute normative practices of ‘good journalism’ is through industry-wide research. In 2014, World Press Photo commissioned a research project on the ‘Integrity of the Image’ in order to assess current practice and accepted standards related to the processing and manipulation of images in documentary photography and photojournalism (Campbell, 2014). The report attempts to identify norms and conventions of “permitted alterations” as they manifest throughout the combination of codes and practices that embody accepted standards (p. 10). As cited in 1.2, Campbell emphasises how digitisation has disrupted the validity of claims for ‘objective truth’. Regardless, “the “credibility of news and documentary photography is conventionally secured in terms of objectivity: the faithful recording of the events and people before the lens is said to secure truth” (p. 3). In order to avoid complete fatalism about the possibility of being able to secure an objective truth, Campbell argues that visual journalism must move “from the ontology of the image to its pragmatics, shifting our concern from what images are to what images *do*” (emphasis in original; p. 16). He describes how as computational photography has unsettled truth-claims to the point that “the idea of objectivity is no longer tenable”, forces practitioners to shift their focus away from what the image is, towards “the purpose of the images, the work of the images, the function of the images, what producers want them to do, and what consumers want them to do” (p. 16).

Deepfakes – at least, in their initial uses – were widely used as tools for deception, manipulation and abuse. Following the first deepfakes that circulated on Reddit, a Vice technology journalist wrote an article entitled ‘AI-Assisted Fake Porn is Here and We’re All Fucked’ (Cole, 2017). As more and more users began to create their own versions of the videos – including the development of an app to allow users without a computer science background to create their own videos (Chawla, 2019) – the journalist followed up with another article, a month later: ‘We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now’ (Cole, 2018). The European Union’s Artificial Intelligence Act proposal – which, at the time of writing, is working its way through EU legislature – includes deepfakes in the scope of its regulation. The provision aims to “protect natural persons from the risks of impersonation or deception when an AI system generates or manipulates image, audio, or video content that appreciably resembles existing persons, place or events and would

falsely appear to a user of this system to be authentic and truthful” (European Commission, 2021). Fernandez (2022), writing on the difficulty of regulating against deepfakes, groups the general consensus among scholars, companies and media outlets on a definition of deepfakes as converging around two elements: firstly, that they entail the use of AI-based technology, and secondly, that they are created with the intent to deceive. However, Fernandez finds, “practical challenges abound” when trying to draw distinctions between deepfakes and ‘cheap fakes’ – particularly regarding incentives for compliance and transparency, alongside enforcement at an EU level. Regardless, engineers, academics and institutions have rushed to create detection methods (van der Nagel, 2020). Across the Atlantic, deepfakes have entered the foreign policy debate, with former US President Donald Trump signing the first federal law specifically addressing the technology in 2019 as part of the National Defence Authorisation Act for Fiscal Year 2020 (Chipman et al., 2019). The focus of the law, however, is solely within the remit of national security – only regulating deepfakes insofar as they are weaponised by foreign actors to target US elections.

In journalistic discourse, deepfakes became synonymous with generalised anxieties around fake news and post-truth futures. Focusing on the 18-month period from January 2018 to July 2019, shortly after a synthesised video of US Speaker of the House Nancy Pelosi circulated on social media, Wahl-Jorgensen and Carlson (2021) analysed journalistic discourses on deepfakes as the future of fake news. Through their analysis, the authors found that the story of deepfakes told through conjectured specificity – that is, specific visions of worst-case scenarios – is of the growing encroachment of ‘bad actors’ facilitated by a changing media ecology, particularly social media; developments that have worked symbiotically with the rise of authoritarian politics and shifts in global power (p. 817).

1.3 Hybrid modalities and Industry 4.0

1.3.1 Synthetic media on screen

Over the past half a decade, synthetic media and deepfakes have been used in films to overcome hurdles in the production process or augment the experience of the final product for audiences, with varying degrees of transparency about the integration of the process. The technology has been employed, for example, when an actor or main character has passed away or is otherwise unavailable. At the height of the COVID-19 pandemic, French soap opera “Plus Belle La Vie” utilised the technology to swap in the face of an actress who was self-isolating onto a body double to not disrupt the storyline (Jackson, 2020). The 2019 Disney+ production of the Star Wars spin-off *The Mandalorian* utilised similar technologies to de-age actor Mark Hamill, to make him appear as a

young Luke Skywalker for a cameo at the end of series. However, the quality of the visual effects were criticised, with the UK's *Radio Times* writing "most agreed he looked a little janky" (Chapman, 2022). After a YouTube creator operating under the username of Shamook made his own version of the scene using his own rendering of Skywalker, which quickly went viral as a more sophisticated and convincing likeness, Lucasfilm, the production company behind the series, hired him to work on visual effects for the sequel, much to the celebration of audiences.

In December 2020, UK broadcaster Channel 4 aired an *Alternative Christmas Message*, a comedy-parody in which a synthesised Queen Elizabeth II thanks Channel 4 for "giving me the opportunity to say whatever I like, without anyone putting words in my mouth" before proceeding to dance across the screen (Bartlett, 2020). The film unsurprisingly triggered a round of debate in the UK, with media regulator Ofcom receiving 354 complaints about the broadcast – which they ultimately decided to not take further action on as the film had a clearly "satirical tone" and left viewers "in no doubt that it was not real" (BBC News, 2021). Regarding the furore, researcher Henry Adjer told *The Guardian*: "As a society, we need to figure out what uses for deepfakes we deem acceptable, and how we can navigate a future where synthetic media is an increasingly big part of our lives" (Blackall, 2020).

Adjer's comments – and viewers' varying reactions to different usages at large – constitute part of the wider conversation on how AI-assisted synthetic images are received. What emerges from the commentary is that the question ultimately comes down to one of intention, qualified by consent of the actors or contributors involved, and disclosure to audiences. In a fictional production – such as *The Mandalorian*, the threshold is higher: when viewers watch the series, they are implicitly aware that they are consenting to watch a visual product that has been highly manipulated with visual effects – indeed, it forms part of the appeal, augmenting the creation of the fictional world and the experience as a whole. While the *Alternative Christmas Message* was not fiction, as such, it was a provocative production designed for entertainment purposes, and made no attempt to obscure the fact that deepfake technologies were used – the last moments of the broadcast panned out to show a studio with a green screen, and the original actor whose face was swapped with that of the Queen was revealed.

Deepfakes – and arguably advanced digital effects overall – hold a more contested place in the realm of documentary filmmaking. The 2021 documentary feature film *Roadrunner*, about the life and death of the chef and author Anthony Bourdain, was embroiled in controversy following revelations that a part of the film used AI-generated audio synthesising Bourdain's voice without disclosing it to its viewers. For three lines in the film, director Morgan Neville commissioned a software company to recreate Bourdain's voice reading an email to a friend. Once news of the

synthesised audio broke, the subject trended on Twitter with heavy outcry. Film critic Sean Burns wrote: “I feel like this tells you all you need to know about the ethics of the people behind this project” (Wolfe, 2021). Speaking about his use of the synthetic audio, Neville told New Yorker correspondent Helen Rosner “we can have a documentary ethics panel about it later” – a comment that, according Rosner, “did not help assure people that he took these matters seriously” (Rosner, 2021). To deepen the controversy, Bourdain’s widow publicly denied having ever given permission for his voice to be synthesised. In conversation with Rosner, *MIT Technology Review* editor Karen Hao speculated that the alarm might also be due to the “hybridisation of reality and unreality” (Rosner, 2021). “It’s not clearly faked, nor is it clearly real, and the fact that it was his actual words just muddles that even more,” Hao said.

A similar ‘muddling’ – of a real speech, performed by a synthesised voice – formed the central tenet of the short film *In Event of Moon Disaster*, which premiered at the IDFA DocLab in Amsterdam in November 2019 in the form of a physical installation recreating a 1960s era American living room. The seven-minute film utilises a real speech prepared at the time of the moon landings by presidential speechwriter William Safire for the then-US President Richard Nixon, in case the mission went wrong, and the astronauts were not able to return to earth (Usher, 2021). Through “a variety of techniques of misinformation [...] from simple deceptive editing to more complex deepfakes technologies,” the seven-minute film created an alternative history to challenge audience to “consider how new technologies can bend, redirect and obfuscate the truth around us” (Panetta & Burgund, 2019). In their review of AI’s implications for documentary media, researchers Kapur and Ansari reflected on the film, arguing that the convincing manipulation of past events “puts under scrutiny the practices of representation and the appeal of AI as a medium. By bringing the mechanics of the process to the fore, the makers set up as viable the analysis and inquiry of (dis/mis)information using (re)imagination.” (Kapur & Ansari, 2022, p. 176).

1.3.2 ‘Smart storytelling’

February 2022 saw *Studies in Documentary Film* – the first peer-reviewed scholarly journal devoted to the history, theory, criticism and practice of documentary film – produce a special issue, *Smart Storytelling* (Schleser, 2022). The special issue is a reaction to the host of emerging digital technologies from a range of actors – from advanced tools from technology giants such as Apple, Meta and Google to more democratised and accessible innovations such as mobile filmmaking. The goal of the issue, as stated in the guest editors’ introductory text, is the exploration “changing creative ecologies” and their “potential for opening storytelling to these new forms and formats and its implications for re-shaping screen-based storytelling practices” (Schleser, 2022, p. 98). The convergence of documentary practice with novel tools and formats – virtual reality, augmented

reality, artificial intelligence and interactive elements – have been termed by scholarship and professionals in the field ‘Industry 4.0’. The issue – which features six articles and three book reviews – is testament to the burgeoning field of scholarly engagement with the ways in which emerging technologies complicate and augment traditional production processes, distribution models and critical understandings of non-fiction film. The text engages in how a whole host of “key developments in computational creativity” have impacted the documentary field. Although most of these are beyond the scope of this paper, the issue provides a framework to understand the ways in which non-fiction film is “embrac[ing] data storytelling and AI as new assets and co-creators in the production process” (Schleser, 2022, p. 102). In the conclusion of the opening text, Schleser writes, “these emerging media forms are now influencing mainstream, media and are thus crucial for scholarly investigation.” He continues: “Creative innovation provides new forms and formats to engage communities and connectivity is thus not only a technical, also social element within storytelling” (Schleser, 2022, p. 108).

In an article on deepfakes and documentary practice in the age of misinformation, communications researcher Craig Hight reflected specifically on the role of *Welcome to Chechnya*. Seeking to present a “counter-narrative” to the climate of alarm over the disruptive potential of synthetic tools, Hight argues that “it is essential to consider the increasing sophistication and potentially existential challenge of deepfake technologies in relation to the array of manipulations which are already inherent to documentary practice” (p. 1-2). Hight argues that documentary media are already increasingly entangled with processes of softwarisation: from non-linear editing, to interactive formats, to reconstruction, all of which “suggest a more complex, performative, constructed and playful range of media than is traditionally assumed of documentary” (p. 8), finding that the use of deepfakes in *Welcome to Chechnya* demonstrates that synthetic media can “clearly operate as a new part of the repertoire for documentary designers” (p. 9).

The case study opens up a wealth of questions. To what extent can deepfakes claim an icon value, or index value? Does the presence of the advanced manipulation itself prohibit such truth-claims in their entirety, or might it instead amplify the question of trust in the producer(s) of such images? How does this change viewers’ relationships to image-creators and distributors, and who are the arbiters of such authority? Must it be negotiated on a case-by-case basis ad infinitum, or will journalistic frameworks of ethics and responsibility suffice? Can a globalised, hyper-digitalised, post-truth environment support such a leap of trust, or are deepfakes already so embedded in the visual language of deception that they will never shake off those binds? Has the visual language of journalistic anonymising already reached its final resting point in blurry ovals or physical disguises, given the saturation of false and deceptive media and the weariness of audiences, or is it a realm with space for innovation? Such are the questions that this study seeks to engage with.

2. Methodology

2.1 Research aims and design

This study aims to answer the research question: *How do practitioners view the benefits and limitations of using deepfake technology to disguise participants' identities in documentary film and video journalism?*

Despite the most developed example of the use of deepfakes for this purpose being in documentary filmmaking, the hypothesis of this study was that such techniques could be a useful tool to video journalists, or anyone who engages in witness filming, particularly when working with at-risk participants. As such, this study included in its scope practitioners who define themselves as journalists, who have worked in the medium of video.

In order to explore practitioners' "thoughts, feelings, or interpretations of [the] meaning and process" (Given, 2008, p. 26) of the use of such tools, this study is qualitative in nature. It aims to produce an "in-depth and interpreted understanding [...] by learning about people's [...] experiences, perspectives and histories" (Ritchie & Lewis, 2003, p. 22). By examining the nuances of practitioners' views on the integration of deepfakes into their practice, the research seeks to outline some of the key concerns, patterns and understandings of this process, to provide initial insights from which more expansive and generalisable research with higher external validity can be conducted.

2.2 Data collection

In order to gain meaningful insights on practitioners' perceptions of the benefits and limitations of the use of deepfakes, criteria for interviewees was that they must currently work, or have worked, in video journalism or documentary filmmaking, and that during this time on at least one occasion they must have worked with sources who have requested anonymity, who they have subsequently visually represented in their work.

This study employed purposive sampling in order to identify and make contact with potential interviewees who would be the most "information-rich" with regards to the research question (Schreier, 2018, p. 88). Beyond the above criteria, in order to engender the highest level of representativity possible, effort was made for sampling to be as heterogenous as possible, in order to enlist interviewees from a diverse range of working environments, experience levels, genders, ethnic backgrounds and countries of residence.

The process began with a series of open calls via social media networks, starting with two private industry forum Facebook groups. The first was ‘The Vulture Club’, a private Facebook group created following the deaths of journalists Tim Hetherington and Chris Hondros in Libya in 2011, which serves as a confidential discussion forum and community platform for freelance journalists, photographers, videographers and fixers globally. The second was ‘GiF Exchange’, a private Facebook group created by the organisation Girls in Film, an international online platform and community designed to facilitate professional connections between women working in the film industry. However, posts calling for interviewees on both platforms garnered a total of two responses – making it necessary to expand the approach. Callouts posted and reshared on social network accounts – Instagram and Twitter – garnered several more participants.

Potential respondents were also researched and directly approached, via emails available on their websites or social media pages. Response rates were low – approximately one response for every five messages sent out. Further, some of those who initially responded subsequently stopped contact, or became unavailable. Some cited being unavailable due to being on assignment in Ukraine or Afghanistan, some gave the justification of illness or personal reasons, others did not provide a reason at all.

The academic enquiry of this study necessitated a broader focus than merely in-depth interviews with the team behind *Welcome to Chechnya* (France, 2020). Much has already been written on the film in film critic circles, online publications and in interviews with the director, David France, as well as, some within academic circles. To avoid the deficit of external validity that a study focused only on the actors involved in the production of this specific film would risk, the purpose of this thesis is to attempt to examine the use of deepfakes and synthetic media *in the context* of other working practices across the industry more broadly to elucidate insights on the future of this technology.

That said, the status of *Welcome to Chechnya* as the only fully-fledged and published example of the use of full-mask deepfakes as a tool for anonymisation made it unavoidable that the film was integral to the assumptions and design of this research project. As such, issues and themes raised by film formed the case study that became the context for the data collection (Flick, 2018, p. 3) The film informed the research proposal, as well as the research design, and featured as an anchor point in the semi-structured interviews with practitioners, through the screening of a short clip from the film to inform the conversation.

For these reasons, the first stage of the data-gathering process constituted an unstructured interview with the visual effects supervisor on the film, Ryan Laney. An unstructured interview

format, in which neither the question nor the answer categories are predetermined (Minichiello, 1992) was chosen to allow for the development of “a more sophisticated understanding” (Murray, 2018, p. 268) of the nuances of Laney’s individual experiences, purposes and reflections of the process of creating deepfakes for this purpose. As this study’s research question is exploratory in nature, the interview with Laney sought to gather descriptive data on the processes behind and experiences of the creation of synthetic masks for the film that would go on to inform the design of the semi-structured interviews of journalists and documentary filmmakers.

The interview with Laney (Appendix) coalesced around certain points:

- Limitations of legacy tools, such as shadows and blurring
- Practitioner-source relations, in terms of standards of efficacy of anonymisation and security in an internet context
- Aesthetics of deepfakes and other post-production methods
- The boundary between truths and lies in visual storytelling
- The importance of context and purpose in any journalistic work
- Financial and resource constraints

These talking points contributed to the final design of the semi-structured interviews with practitioners. A semi-structured interview design was chosen for the practitioner data collection, in order to ensure that the research question was adequately addressed, and “the same topics form the basis for questioning” (Roulston & Choi, 2018, p. 233), while creating flexibility and space for practitioners to elaborate on their perspectives and experiences.

Care was taken in the sampling process to gather as broad a cross-section of experiences and views as was feasible within the timeframe of the study. Of the eight practitioners interviewed, five were women and three were men. Their main working regions covered several continents: Asia, Africa, North America, South America and Europe. They also had varying levels of experience, all having worked in the industry for between one and 15 years. The practitioners had worked with a wide variety of clients and distributors with varying models of funding, format and editorial oversight.

Anonymity was offered to all interviewees, with some preferring not to be identified in the research – as such, for the sake of uniformity, all participants except Laney are pseudonymised henceforth with their initials. Table 1 below provides a brief overview of the practitioners with which semi-structured interviews were conducted.

Pseudonym	Main working regions	Gender	Primary mode(s) of work	Primary client(s) and/or distributor(s)
AM	Brazil, Denmark	F	Photojournalism, documentary filmmaking	Al Jazeera, BBC Brazil
AD	USA, South Africa	M	Online news, broadcast news, video reportage	Bloomberg TV
AJ	India	M	Online news, video reportage	The Caravan
DT	Greece, Canada	F	Photojournalism, documentary filmmaking, video reportage	Vice News, PBS, BBC
JM	Japan	F	Online news, broadcast news, video reportage	Nippon TV
KF	Indonesia	F	Documentary filmmaking	Daily Guardian
LC	Hong Kong, Central African Republic	F	Documentary filmmaking, video reportage	Vice News, National Geographic
TA	Honduras	M	Photojournalism, documentary filmmaking	National Geographic, The New York Times

Table 1: overview of practitioners

The interviews were semi-structured: key questions were asked in the “same way” across each interview, with some probing for follow-up information depending on the practitioners’ answers (Ritchie & Lewis, 2003, p. 111). All interviews were conducted on Zoom, by necessity: practitioners were interviewed from multiple continents, and funding restrictions would not allow travel to conduct interviews in person. Accordingly, almost all practitioners cited having busy schedules. Online interviews, therefore, were much more accessible, convenient and time-saving (Gray et al., 2020, p. 1292).

The first part of the interview asked sources to describe their career background: the type of visual media they’d worked in, how long for, for which clients, and how they would describe their

style and target audiences. The second part was regarding previous experiences with source anonymisation: the number of times the interviewee had worked with sources who'd requested anonymity, the information excluded or included, the in-camera or post-production tools available to them at the time and which they elected to use, or didn't, the editorial oversight enacted upon them, and their feelings about the options available to them for anonymising sources.

At this point of the interview, a visual prompt was shown to each interviewee – a clip from the film *Welcome to Chechnya*, in which the main character, 'Grisha', who is digitally masked up until this point, reveals his true identity – Maxim Lupanov – at a press conference reciting his experiences (1:22:18 to 1:26:46). It is at this point of the film that the mask 'dissolves' on screen, and Lupanov's real face is revealed. This part of the film was chosen as the visual prompt as it neatly illustrates the contrast between the synthesised face and the real face. Before screening the clip, a brief contextualisation was given: of the synopsis of the film, the events the interviewee was about to watch, and the fact that the first part of the clip was visually manipulated through AI-assisted technologies.

Following the screening of the visual prompt, the final part of the interview consisted of a discussion to interviewee's reaction to the film, in which they were asked whether they could imagine in a situation in which they would use full digital masks, or deepfakes, such as these, what they could imagine their concerns might be throughout the process, and how they would feel if the use of such technologies became more commonplace within documentary filmmaking and video journalism. The video recordings of the interviews were transcribed manually.

2.3 Data analysis

As the goal of this study was primarily explorative in nature, thematic analysis was employed due to its flexibility. As the use of synthetic media used for the purpose of anonymisation is not a well-established phenomenon within journalism, a key priority of the research was to avoid superimposing a strict theoretical framework onto the data collected, and instead, working inductively *with* the data to draw out patterns of meaning. As such, the analytic process began on the semantic level – a 'bottom-up' style of analysis devolved from theoretical assumptions. Following the construction of initial themes, the themes were reviewed and refined in combination with theoretical insights to enable richer and deeper final analysis of how the interviewees' perceptions of deepfake technology interact with theoretical understandings of how practitioner's view the iconic and indexical value of synthetic images and how these understandings are operationalised when combined with each practitioner's ethical approach to both their work and their engagement with the video journalism and documentary filmmaking industries at large.

The process of familiarising with the data – the first phase of Braun and Clarke's (2006) six-phase method of conducting thematic analysis – was achieved in part through the interactive means of collecting data (semi-structured interviews), then consolidated by manual transcription of the recordings, and another close reading of the data corpus once the interviews were complete. The aim of the second phase, the generation of initial codes, was to identify the most basic segments of the data (Boyatzis, 1998, p. 63) to form the building blocks for the later stages of interpretation to build themes. From these initial codes, the initial thematic map was developed to illuminate preliminary connections between codes and identify patterns of meaning – phase three in Braun and Clarke's framework. In the process of constructing this map, codes were loosely categorised in terms of their relation to four patterns, or candidate themes. A large proportion of the initial codes – constructed as they were at this stage, as the simplest building blocks – related to multiple themes, as they represented an element that was discussed in some capacity in relation to different stages or aspects of the production and distribution process. After reviewing these four candidate themes using Patton's (2015) dual criteria for assessing categories: internal homogeneity and external heterogeneity, the data was refined and re-coded to create four final themes and sub-themes.

2.4 Refinement of initial themes

The initial coding process resulted in the generation of 31 codes (Table 2) with their respective frequencies. The codes are loose topics or talking points that were referenced in their most basic form throughout the interviews. The frequency, whilst indicative of patterns of reference, does not necessarily indicate importance with regards to the research question (Braun & Clarke, 2006). The salience and significance of each code was interpreted through an assessment of the whole data corpus.

Code label	Frequency
aesthetic value of deepfakes	8
audience perceptions (other)	5
audience's trust	10
criminal associations	2
decide approach with team	1
deepfakes as tools of deception	8
dependent on product type	3
differences in global south context	5
editorial oversight	9

fiction v reportage	10
finding the balance	21
flexibility	3
humanising	3
informed consent of audiences	3
informed consent of sources	16
limitations in available tools	1
narrative choices (other)	11
novelty of technology	17
other identifying features	3
practitioners' ethical responsibilities to sources	23
preference: in camera	9
resource restrictions	16
security benefits	6
security in an internet context	9
slippery slope	7
source's motivations	2
source's trust in journalists	5
thinking mask is a real face	2
unintended consequences	14
visual language of anonymity	31

Table 2: codes and frequencies

Connections and links between these codes were subsequently visualised in the form of an initial thematic map (Figure 1). The process of construction of this map resulted in the formation of four initial themes, or patterns of meaning. These were the key factors that practitioners' concerns around benefits and limitations gravitated around and illustrated four stages of considerations that practitioners make while choosing the tools that they will integrate into their work: the protection and navigating of their relationship to sources, practical considerations that govern the feasibility of their work, the final aesthetic output of their work, and how their work plays into industry dynamics at large. Beyond these four main themes, each theme was further divided into two to three sub-themes, talking points or referents that formed the basis of practitioners' perceptions. These are illustrated in the secondary thematic map (Figure 2).

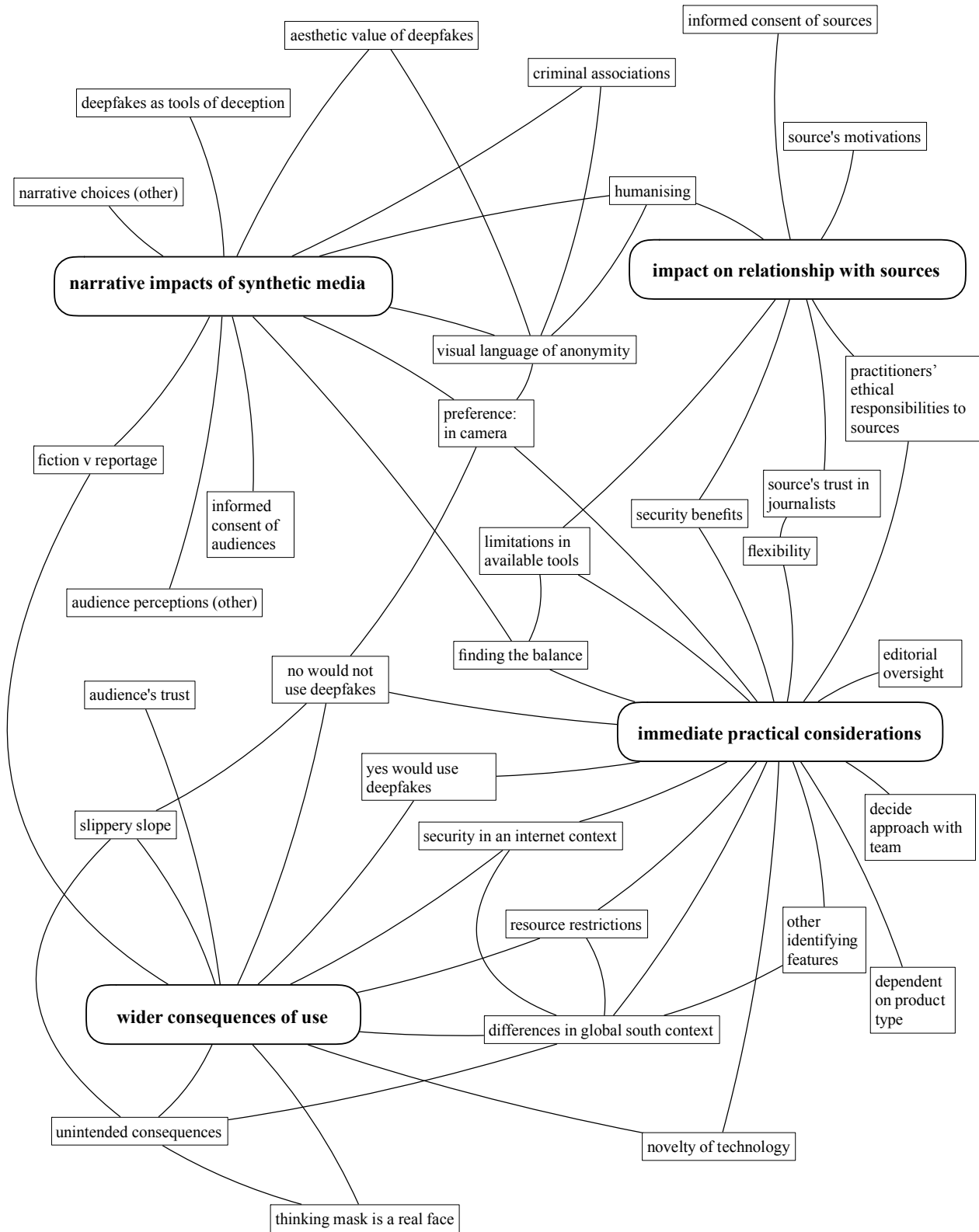


Figure 1: Initial thematic map

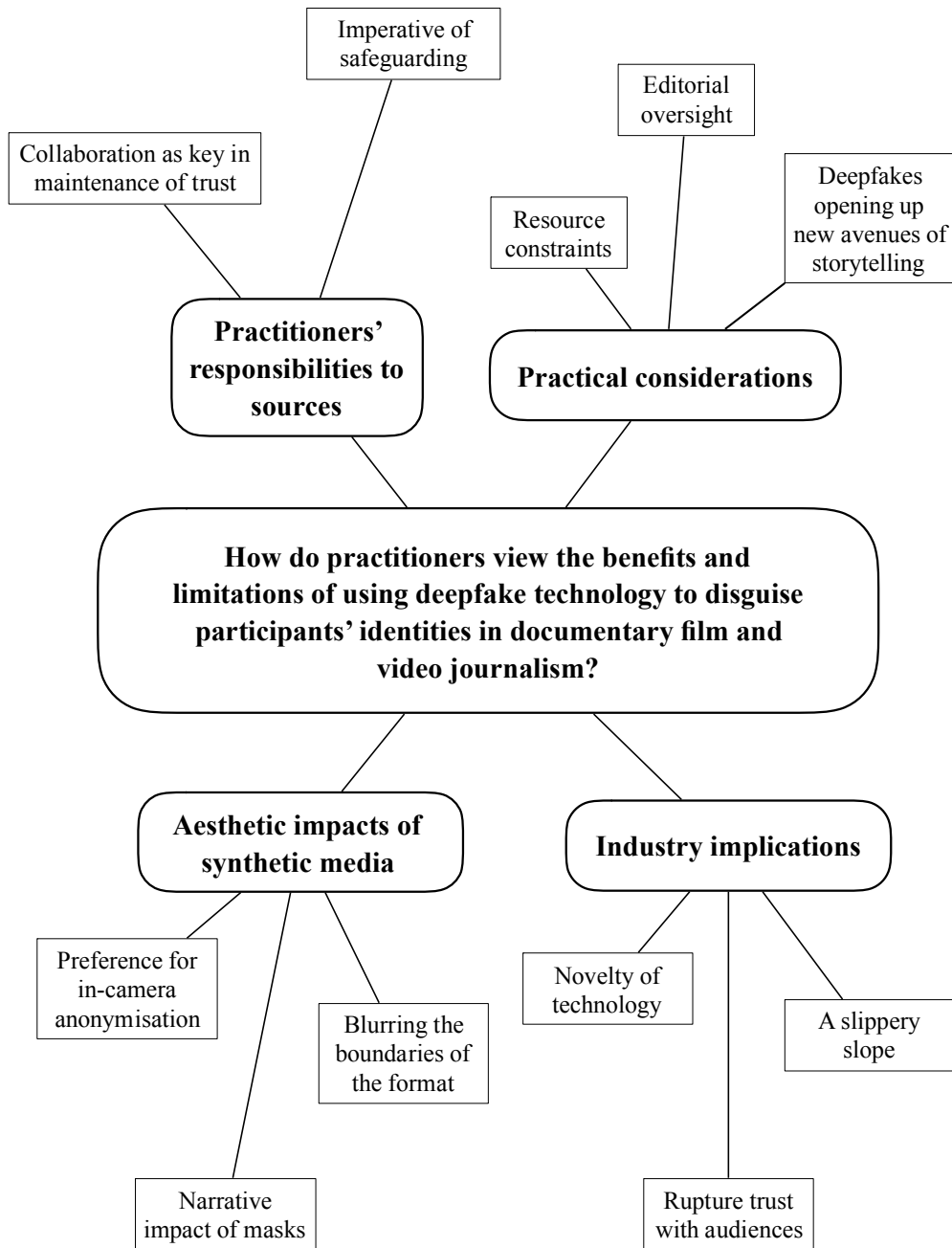


Figure 2: Secondary thematic map

3. Findings and Discussion

Practitioners' perceptions of the benefits and limitations of using deepfake-like technology to disguise participants' identities in documentary film and video journalism were expressed along the lines of four key themes. The first was how practitioners conceived of their responsibilities to the sources they work with: how collaboration formed the cornerstone of trust, which was key to their working relationship, and the practitioners' role in protecting the safety of the source. The second was the practical considerations at play in their working practice: a key sub-theme was discussion of resource constraints, and how this might affect the use of synthetic media to anonymise, alongside the acknowledgement of how editorial decisions are influenced by supervisors, distributors or clients. Also raised as a practical consideration was a positive: that having access to an expanded toolkit that involved the ability to create deepfakes might open up previously unavailable avenues for storytelling. The third theme, or key consideration, was the aesthetic impacts of blending such highly digitised images into their work. Many expressed a preference for in-camera tools, as opposed to post-production manipulation, to disguise sources, linking this to normative understandings of the role of documentary storytelling and a resistance to excessive manipulation. Also questioned was the narrative impact of masks, and how such factors impact the overall visual appeal of the work. The fourth theme was the broader industry implications of greater use of deepfakes as a tool for practitioners. Within this theme, practitioners addressed the novelty of the technology, and the correspondent concerns that raises – that at the stage a new tool is brought out into the world, the full implications cannot be known, and it might be the beginning of a 'slippery slope'. Practitioners also raised that the use of deepfakes, however well intentioned, might further rupture trust between audiences and practitioners on the credibility of their work. The first section of this chapter (3.1) overviews in detail how respondents differentiated in how they discussed these four themes; the second section (3.2) analyses these responses in relation to theoretical understandings of the image and correspondent ethical practice, alongside a discussion of the limitations and implications of this study.

3.1 Findings

3.1.1 Practitioners' responsibilities to sources

Without exception, all practitioners emphasised that the protection of the source who requested anonymity was paramount, with their sense of responsibility towards sources and

frameworks for navigating that relationship informing how they perceived the benefits or limitations to different tools of anonymisation. All practitioners demonstrated an implicit awareness of the large amount of identifying information conveyed through visual imagery, and such, their responsibilities in hiding or removing information that could at a later stage materially harm the source. Within accounts of responsibilities towards sources, two sub-themes emerged: localised responsibilities to the individual in which the journalist was interacting with and portraying (3.1.1.1) and wider responsibilities that include an awareness of possible repercussions following the publication and distribution of the documentary or video report (3.1.1.2).

3.1.1.1 Collaboration as key in maintenance of trust

Practitioners described the ongoing dialogue between sources and journalists as a collaborative process, with sources having the final say in how they are depicted. Within this understanding, practitioners differed on their perceptions of their responsibility as an individual journalist. Those working for larger broadcasters attributed much of the trust in handling sensitive information to the wider reputations of their companies, demonstrating deference to supervisors' editorial guidelines for when and how to anonymise their sources, where practitioners who worked primarily as freelancers tended to emphasise at greater length their *personal* responsibility to make the right judgements of which information to include or exclude, as individual journalists.

AD and AJ, recalling experiences working with US- and India-based media outlets, both emphasised that while sources were in control of what part of their identity could be included, the institutions they were employed by had editorial guidelines that to some extent dictated the options available. AJ's organisation stipulated that sources could only be anonymous if revealing their identity could affect their life or livelihood in some way, and his supervisors had an in-house preference for silhouettes. Likewise, requests for anonymity at AD's organisation entailed "a little more background checking [...] than other stories", and such requests are rare, as it is only in a handful of circumstances that people have "a legitimate concern of being materially influenced or impacted by speaking to the media." AD also emphasised that in his experience, trust in his news organisation was high – meaning that if anonymity was granted, sources implicitly trusted journalists in responsibly distributing the information they provided. JM, a long-time correspondent for Japanese broadcaster Nippon TV, reported that in her experience, sources also placed a high level of trust in her organisation's handling of their safety, attributing that fact less to the reputation of her specific organisation, but Japanese media culture as a whole.

Interviewees who primarily work as freelancers – who are therefore less likely to be working under the mantle of a media house, or commonly working for multiple clients with different editorial approaches – heavily emphasised the role of the individual journalist in decision-making processes for how to anonymise vulnerable participants, as well as the building and maintenance of trust. A common motif was flexibility. KF, who has produced independent social issue-led short films for charities, said “I’m not so persistent, as long as me and the subject are comfortable with [whatever method of anonymisation we choose]”. AM, also recounting an experience of working on a social issue-led short film, with survivors of sexual violence, said she began the process by anticipating that they would most likely not want to be on camera, describing how she combined the use of silhouettes with actors’ voices and faces reading a script – she agreed that ultimately, all choices were source-led.

TA, who has worked extensively in Honduras with members of organised crime groups and victims of drug-related violence, described the transactional element of the collaborative source-journalist relationship. “Because they’re speaking with you, they want to tell their story,” he said. “Most people are not doing you a favour by doing this. They want something out of it.” He also described how the trust sources placed in him as a journalist varied: “so a lot of the time we were with gang members and I mean, I didn’t have the ultimate say then [...] the gang members [...] just used a mask or something, like they would physically use a mask or tie a T-shirt round their face and if you don’t like it, well, go fuck yourself [...] so, not giving us [the filmmakers] the benefit of the doubt.”

3.1.1.2 Imperative of safeguarding

Implicit in all interviews was the imperative of safeguarding, and the importance of the journalists’ role in that, however two practitioners – DT and TA – gave explicit examples of repercussions that occurred, or could very easily occur, should too much identifying information be revealed. Recounting her experiences filming with unaccompanied refugee minors in Greece, DT emphasised the role of the journalist in drawing the line of what information to include – beyond what the source requested. “They didn’t know, they didn’t care” she said, recalling filming interviews with young boys who were homeless, earning money by offering sex acts for money with older Greek men. “They weren’t even aware of the issue, they were like, fine, I don’t mind showing my face. But we were like, no, we’re not going to show your face, and we’re going to change your name [...] they might become a target and we have to protect them.”

TA recounted another example of repercussions of disclosing too much information – this time, in the context of the consequences if a production team fails to honour a source’s request. He

recalled a film project which a team of British journalists came to Honduras looking to make a film about drug violence and speak to gang members. TA described how a friend and colleague of his was a fixer on the project – he had coordinated with high-level gang members and made assurances to the sources that they would remain anonymous, before bringing the British crew to do the interviews, which they did. “When it came out, and to this day, I don’t get it, even though it was communicated to [the production team], they blurred out everyone’s faces, except the bad guy, the most bad guy,” TA said, becoming heated as he described it. “So, they got the video on their hands. And it was a disaster. Immediately, like, without a second’s thought they went after my friend, very nearly killed his whole family [...] I helped him leave the country because I knew people at the Rory Peck Trust [an organisation for the protection of journalists]. It’s the most malfeasance that I’ve ever heard anyone do [...] it’s absolutely criminal,” TA said. “And this is the importance of doing like this privacy anonymising [...] because it isn’t just with victims, like it’s also with people who are extremely dangerous and will take repercussions not with the star journalist production company that comes from the North Atlantic, but with the fixers that are left behind,” he said. He continued: “It’s going to the point of responsibility, you can’t leave anything to chance [...] it’s you, you as a journalist, you have the final say. You’re the one designing everything. So you must abide by the requests of the people you’re trying to portray, and you must also sometimes take further steps if there is even the slightest possibility that there might be blowback because ultimately everyone is going to have access to whatever you’re making.” Continuing in reference to broader examples, TA described how he sees the process as “not even a negotiation [...] I don’t know if this is the benefit of being independent – but it’s a rule that just cannot be... it’s sacrosanct, you know? It’s a conversation you often don’t even need to have, you can see the person’s fears in their eyes.” He described how the images he produces are created in dialogue with the sources he works with: “I ask: what do you think? What if I take a picture of your necklace? Or your silhouette behind – not even the silhouette – your vague shape?” He emphasised how ultimately “it depends on how much the person is willing to show [...] then having enough plausible deniability, or plausible ambiguity rather.”

TA also expressed concern that the use of deepfakes would not be effective to the goal of anonymisation, due to all the extraneous detail – beyond the face – communicated through video. “My first reaction to it is the fear that locations can be identified”, said TA. “There’s a matter of resources, when the people you’re hiding from have nice infinite resources, then it’s not really that hard to figure out places with pictures and video [...] and so like, if you slip up once, great, you had that fancy deepfake technology, but you didn’t mask the exterior, which I would assume [the producers of *Welcome to Chechnya*] didn’t.”

3.1.2 Practical considerations

The second theme that emerged throughout interviews with practitioners on the benefits and limitations of synthetic media was the practical considerations that might arise. Predominantly, discussion of practical matters was presented as a concern, or limitation, though two expressed that access to synthetic tools could allow for greater access into stories that would not have been able to be covered using legacy tools, such as blurring (3.1.2.3). Resource restrictions (3.1.2.1) – both time and budget constraints – were highlighted in both documentary and journalism but were emphasised as an especially prescient concern in the context of producing reportage on a fast turnaround. Another practical consideration raised was editorial oversight (3.1.2.2) especially by those working under the mantle of bigger media organisations, who have their own frameworks and guidelines for anonymisation processes, but also for freelancers, in terms of the distributors.

3.1.2.1 Resource constraints

One of the most salient responses from interviewees was concerns over the resources that introducing synthetic media into their work might entail, primarily in time and money. Resource restrictions were raised obliquely within every interview, with some identifying it as their key concern: “My first reaction was how much did this cost? And how much time? [...] it’s just an extra challenge, a big challenge” said DT, a freelance journalist. She highlighted that in observational-style documentaries, in her experience resources are already usually stretched in terms of collecting the footage, following someone over a prolonged period of time. “But then, you know, having to actually digitally make this mask [...] it’s time consuming, and I guess a lot of money to do so. Which normally, we don’t have in documentaries.”

Another, JM, who worked at a Japanese broadcaster, highlighted the time such a process would take as being the main factor in whether it stood a chance at being integrated into newsroom practices. “So, for example [...] normally we decide what to report in the morning, and take this film in the afternoon, and it’s on in the evening. So, we really have limited time. So, if this is like one click or like – no, no, maybe if it takes less than ten minutes [...] based on our workflow, I think we can’t use it for daily news.” She then highlighted that it would be a possibility to be used as part of a package – a five-minute mini-documentary within the broadcast – which reporters sometimes have a lead time of one to two weeks to produce. “If we do that, we could spend like, maybe a half day to do this [...] it could be really valuable. But still, like the timeline is very short, because we also have to cover daily news.”

One respondent, AM, framed the question of resources in the context of the geographic environments in which she has worked, highlighting that filmmakers and journalists in the Global South, in particular, struggle to access funds and resources. “I just don’t think that we would get this technology easily, you know, like, because sometimes we have grants that are, I don’t know, \$5,000 for a feature documentary or even less [...] I think that this alone would take more than the entire budget.”

3.1.2.2 Editorial oversight

Five respondents identified editorial oversight, clients or distributors as a contributing factor as to whether they would ever use deepfakes to anonymise their sources. One, LC – primarily a freelancer – said that even when not operating under the framework of a supervisor as such, distributors would be “perhaps hesitant” to take on a project using such a technology. Two, AM and AD, who worked in contracted positions at news companies expressed that the decision-making processes around modes of anonymisation were explicitly guided by editorial frameworks decided by supervisors or company boards, both implying that that would be a restriction. “I’ve always had supervisors guiding me or at least their input,” said AD. “Whether a news organisation, like the one I work for, would allow something like that is probably something that would be up for debate [...] certainly, the devices that I’ve used [in the past], blurring, adjusting someone’s voice, there’s a style guide, and there’s a procedure of doing these things, and I wasn’t the first to do that.” AJ also referenced editorial guidelines as a factor: “we had a policy even not to use pseudonyms [...] it was just sort of a house policy, editorial policy in terms of, you can’t ascribe a fake identity, because any sort of name or identity has its own meaning and sort of context that is built into it. So, we would not use a deepfake because that is the visual equivalent of a pseudonym.” AD also reflected on the process of the use of something like a deepfake being approved by supervisors: “I’d imagine that there’d be a whole host of ethical considerations being made at the top level of news gathering before something like would be implemented.”

3.1.2.3 Deepfakes opening up new avenues of storytelling

Two, however, expressed that having a such a tool at hand to anonymise sources could be a bonus to their work, as it could open up access to storylines that they otherwise would not be able to cover in the same way. “Sometimes, we had stories dropped, because we couldn’t show the faces”, DT explained. “If it’s a compelling story, and you need to show not just a part of the interview, but

you need to follow the character a bit, show [them] doing something else and you don't have a face [...] you have to make compromises and see whether it's then worth filming this person or not." She described an instance in which her team was told by a distributor that they had to find more characters when one of their sources said she didn't want to be identified; meaning they used her less on screen, and had to make up the time with someone else that consented to showing their face. "Because it does really make things more difficult," DT said. Explaining the benefits of a tool that managed to preserve anonymity while communicating expression, DT said: "Because you're working with the image and working with the person telling the story. So, you want to see this person's expression, this person's eyes [...] I don't like to have a blur of this person's face."

AM also expressed enthusiasm about the types of stories that could be covered with the added tool of being able to create a fake, but lifelike, face for the interviewee. "I think this would be amazing, game changing", she said of the possibility to superimpose a face. "You could gain access to a lot of places and a lot of things that you normally wouldn't."

3.1.3 Aesthetic impacts of synthetic media

The largest divergence between interviewees centred around their perceptions of deepfakes and synthetic media and how it would – or would not – fit into the end visual product of their work, with the most dominant pattern being strong reservations about the prospect of including AI-assisted images such as deepfakes into their work. Many practitioners expressed a preference for in-camera anonymisation (3.1.3.1) for aesthetic reasons, with others expressing resistance to utilising highly computerised AI-assisted tools on the basis that they challenged principles of objectivity and in some way contradicted the purpose of video journalism and documentary – and that such tools belonged more in the realm of docudrama or other blended formats (3.1.3.2). Practitioners added that the ways in which a face is masked – and the fact that a face has to be masked at all – has a narrative impact in and of itself, which is factored into decisions on which tools to use (3.1.3.3).

3.1.3.1 Preference for in-camera anonymisation

Commonly expressed was a preference for in-camera anonymisation and expression of hesitancy even towards commonly available tools such as blurring, computer-imposed shadows or voice manipulation. "Mainly I do in-camera," KF said. LC said: "It's not aesthetically pleasing [...] usually I will shoot strategically or use lighting in order to obscure faces or ask the subject to cover their face with a scarf and/or hat. I haven't altered voices before." DT made the same point, saying:

“It’s much better to creatively film in a way that you don’t have to do that in post-production.” She explained how she prefers to create a deep depth of field, framing the camera in such a way that an item in the foreground – a plant, for example – remains in focus, and the subject in the background is blurred to the point of being unintelligible beyond their basic form. “It’s more cinematic in a way,” she said. JM said that while she has access to post-production tools, she has not used them in her nine years working as a broadcast. She described how she usually films the back of the head, as well as hands, or the back of the person, with her – as the interviewer – in focus.

AJ also emphasised his preference for avoiding post-production tools, and working chiefly in-camera, mostly with strongly backlit subjects so that their distinguishing features wouldn’t be able to be perceived by the viewer. “In terms of information about the people themselves, I think silhouettes are just sort of the neatest and most compelling way to do it,” he said. “Like you could do things like I don’t know, the back of the person’s head or sort of facing a different direction, or shoot some different angles, shoot the hand or something. But I think it just visually, it doesn’t, it doesn’t feel as compelling if that’s what the viewer has to see.” He described how the purpose of post-production would be mainly to correct for any errors that happened through the filming process – “if they reveal some part of their identity or something like that, then you’d increase the contrast or edit that out,” he said. AM also only used post-production to tweak what was already recorded in-camera, including when recording the actors, so that the visual output was homogenous: “just colour correction on the silhouettes.” AJ added: “There is also generally a sort of a sense of intrigue and secrecy and I think [...] if it’s a silhouette, I think you can sort of sell it better, if they see the person and sort of hear the fuel in their voice. As opposed to like, if it’s in a written text, you’ll be like, so this person has to be anonymous. Readers might be like, why?”

3.1.3.2 Blurring the boundaries of the format

Opposition to post-production techniques tended to be qualified by a belief that the overt, or excessive, use of computer manipulation in some ways contravenes the purpose of video journalism or documentary. “Just like, in an aesthetic sense,” TA said. “If you speak with older cats that have been doing journalism for a long time, they tend to be very orthodox in their trade... [almost] dogmatic in terms of the truth, and worship, like, the objective truth, or as close as they can get to it.” DT made a similar point. “In theory, in documentary, you’re not making that much of a fictitious choice or process, you know [...] you’re just documenting something or someone as it unfolds [...] for me, the documentary is trying to show things as close to what they are. If it’s nice or not nice or

ugly, or whatever [...] maybe tweak the contrast a bit or the colours, of course, you film in a way and then you do some colour correction. But that's different than changing the whole thing."

A point frequently raised by interviewees was that synthetic media is a valid choice – but in utilising such techniques, the filmmaker is moving away from the traditionally-identified documentary format and towards another genre. "I think it's like using an actor," said JM. "Where do you draw the line?" asked TA. "Between crafted fictional storytelling and documentary reportage?" AM said that she does not view "the technology, per se, as an issue", utilisation of such techniques "goes to the border between documentary and fiction." "I think it could be an interesting discussion," she said. "In a way, what you're doing [...] is you're putting a mask and a different face, but it's still a true story." She added that it feels like a different version of a docudrama; in essence, a digitised version of a re-enactment, in which you have an actor say what the source was going to say: "instead of digitally lending a face, you're literally lending a face and a voice to tell someone's story."

3.1.3.3 Narrative impact of masks

AD said that the usage would depend very much on the end format – and as such, in his work in broadcast journalism and short packages, he perceived it to be unlikely that he would use it. "I can see it in the context of documentary," he said. "Perhaps it's a better narrative device within the structure of a documentary than it would be in a one-minute TV news package, which might confuse the news consumer because you're consuming news in a different way. And it might be considered misleading," he said.

Reflecting on the visual impact of anonymising a source, DT said: "We even saw this when we were made wearing the masks during the pandemic. When you had interviews with the masks, even then it was weird, having people not seeing their mouth. But we knew it was a pandemic, we've kind of accepted it. But still [...] sometimes it takes away the credibility in a way [...] Basically, you really want to [...] see who this person is, and it be filmed in a creative way – that [even if you cannot see their face] it's beautifully made, visually nice, and not just blurred at the post-production stage. I don't like it, to be honest. I don't like to just have a blur of this person's face. Or it can work both ways," she continued. "Because you might say, okay, this person is really worried. So maybe it's more credible." Anonymisation lending credibility to the story was echoed by TA, who said "it's those moments that kind of add to the bits that compound the narrative by showing the fear or the apprehension of speaking to someone. You know, everyone knows how people are anonymized. You know, even if you don't work in journalism."

LC added that she always prefers to show real faces, but when anonymisation is necessary, she tries to “incorporate the anonymity into the mood/atmosphere of the image, as it emphasises what is at stake for the subject.”

3.1.4 Industry implications

The fourth theme that emerged from the interviews was concern over the broader industry implications of the mainstreaming of such a tool. Many of these concerns stemmed from a place of not knowing what these might be, due to the novelty of the technology (3.1.4.1), and concern that, regardless of intention, the use of deepfakes would further rupture audience’s trust in journalism (3.1.4.2). Practitioners also expressed apprehension that opening the door to even more use of deepfakes – even for noble reasons – could invite unintended consequences and be a slippery slope (3.1.4.3).

3.1.4.1 Novelty of technology

The novelty of the technology – and corresponding ethical concerns about its misuse – was the most common reason interviewees cited for not wanting to use deepfakes in their work. From an institutional point of view, AD emphasised that while it might be possible, there would be a series of processes and discussions that would have to happen first at the top-level of news production. “I have not seen this being used by any of my peers or the news organisations or content I consume. It's an interesting question that you pose. And obviously, I don't think there's a right or wrong as to whether you should or shouldn't.” However, he added that it would likely test the relationship between audiences and news distributors. “As a consumer of news, whenever I see a source like that, you know, it immediately triggers the suspicion, can this source really be trusted. Does this person have pure motives? Why would this person not be willing to show his or her face or voice? But, you know, this is why superiors have always needed to weigh in in situations like this [to decide whether the source is trustworthy].” He said that if available, it would be “something under consideration [that] superiors around the table would debate. I think it's an interesting discourse.”

TA also voiced his concern that the exciting novelty of the technology would detract from the core of the narrative, making the story being told secondary to the means by which it is delivered. “I'm saying this because the very first thing that shows up on a Google when you search [for the documentary *Welcome to Chechnya*] is all people [asking]: Is it real? And then, you have, you know,

like questions about the actual characters. But the very first thing that I have seen is all about the veracity, on account of the deepfake.” TA said that before you even get to misuse of the technology by third actors, “a lot of journalists are in love with their story, or are very much in love with themselves,” and there is a tendency to forget that the work does not just exist in a bubble, but forms part of the fabric of everyday public life. “You can't fall in love with a technology [...] much like in the early days of being in love with someone you overlook the fact that they don't flush when they go to the toilet, you just think: I love them, I love them. Like the rose-tinted glasses or whatever. And in the project the audience might not have rose-tinted glasses at all and might just see it as bullshit.” TA said: “With such a new technology, we can't think of like the new problems or blowbacks until it's happened, like the example that I told you. It's so clear that you mask, sensor everyone. But these fucking idiots at this channel decided to not censor, the main guy - you would think is obviously like, censor everything. But these guys didn't.”

3.1.4.2 Rupture trust with audiences

TA emphasised that he did not believe that the use of deepfakes in this context would become a widespread phenomenon, instead arguing that they will remain “a curio, an oddity, a selling point.” “Even though it might be used for very specific cases,” he said, “I think it ultimately because of the nature of journalism, and the foundation of trust, given the current climate it will not become a mainstay. It will stay as [...] the thing that's used as the selling point or the focal point, as opposed to just like voice warping, which doesn't cause too much attention to itself beyond the moment that its used.” “The whole foundational pillar of being a journalist is trust,” he said, explaining why he didn't think it would become more commonplace within the industry. “Not only trust with the people you're speaking to, but the general public. Like what if this technology becomes [...] accessible to anyone? What's next? Already people are just like, reflexively saying fake news, or this is staged, or these are crisis actors or whatever, or oh, that's a deepfake. So, you can't just think of it as a tool to apply your trade. You also have to consider what the general public, in the current state of hyper paranoia and polarization, at least in the West.” AJ, speaking on his experience working in an Indian newsroom, voiced a similar point. “That's scary to me, the potential of it in today's era where [...] it's so easy to create fake news to spread some sort of narrative or propaganda. And you see so much of it in India, where there is so much of it used to create anti-Muslim sentiment. And then if it becomes a mainstream part of newsroom, I can very clearly identify that some people would be very willing to exploit that to further that agenda. I hope it doesn't happen. I hope it doesn't,” he said.

JM and KF, however, both emphasised that the disclosure to audiences is key. “I think it's okay to use this skill. If you are open about it, it's really okay,” JM said.

3.1.4.3 A slippery slope

DT emphasised that the use of the technology in this context opens up a whole host of potential future concerns. “As soon as many people start using it, then [...] how do you control it? Where do you stop? And how do you differentiate from you know, truth, fiction, artificial fakes or intelligence or whatever, from what is actually real? What are you really, really seeing then?” she said. Again, she urged the need for contextualisation. “Because already, we have an issue with fake news and disinformation [...] it can make this role [a journalist] even more difficult. I don't think it should become the norm.” As well as legitimate uses of the technology, bringing it into the industry as a common tool concerned DT. While acknowledging that she views the technology as an interesting experiment with boundaries, “it might also become, I don't know, trend or something, just you know, if it's that easy to just kind of switch someone's face with someone else's.” She said that if sources became aware that this extent of manipulation is an option, journalists might receive more requests for superficial changes, such as wanting better skin.

KF echoed the same point, saying that she views it as becoming fashionable within the filmmaking world, citing the expanded use of drones as an example. “It's not always necessary to use drone but since its trend, and it's more affordable right now, when I watch films now, there's always drone, drone, drone, drone – and actually it's not always necessary,” she said. “Or like a stabilizer or something. So sometimes, it has a certain purpose at the beginning or an intention, but especially with this AI system, I don't know – yeah, it might be chaos.” Speaking specifically from her experience in Indonesian media, KF said that, even if originally used responsibly, it is very common for clips to be cut, recontextualised and repurposed online. The risk with AI is that if faces are swapped, or an entirely new face is generated, the journalist could very quickly lose control of the repercussions. TA added: “As opposed to say, like mosaic, or digital masking of other kinds: it can be a slippery slope [...] and it could lend itself to being overused. Sometimes. Obviously, this is not always the case. But just in my small country, working on a very specific subject, I can't tell you the amount of ethical backslides people are so willing to take.” TA said that had he been younger, or less experienced, he might have been more optimistic about the availability of such a tool. “That's why experience is great,” he said. “I mean, I would think that I would fall into that camp, had I not seen the interminable examples of just so much malpractice and irresponsibility with the tools that exist now,” he said. LC

also said that she sees the technology as “over the top” and in all honesty would be unlikely to use it. “It poses novel and interesting ethical questions that I’m not ready to deal with,” she said.

3.2 Discussion

Through these four themes, practitioners expressed nuanced and varied views on the benefits and limitations of using deepfake technology to anonymise sources. In summary, the benefits cited by practitioners of such a use of deepfakes were few. One practitioner, AM, said that the technology had the potential to be “game-changing”, and enable journalists to access and tell stories that they would not have otherwise been able to, a sentiment that was also expressed by DT. Another, JM, said that it could be “really valuable”, although with a caveat that in most circumstances, this tool would not be appropriate or practical. Such comments were qualified by understandings that ultimately, decisions on how, when and who to anonymise would be made in collaboration with the source themselves, with the individual journalist bearing the brunt of the responsibility for their safeguarding.

Crucially, however, much more dominant than discussions on the benefits was discussion on limitations: from practical considerations, heavily influenced by tight resources, to aesthetic concerns, to, especially, the wider consequences across the industry for the integration of synthetic media – both from a place of ‘trustworthiness’ and the possibility of unintended consequences creating a ‘slippery slope’. While *Welcome to Chechnya* (France, 2020) was lauded as a revolutionary use of technology in this context, this study suggests that practitioners’ perceptions of journalistic methods still adhere very much to norms of ‘truth-telling’. As such, while the camera-as-computer has shaken ontological understandings of the indexical value of images (Bate, 2013; Campbell, 2014; Lister, 2013; W. J. Mitchell, 1994), practitioners largely expressed that adherence to the *pursuit* of indexical claims remains foundational to their work, which plays out in attitudes to post-processing tools.

Excessive manipulation was largely problematised: no practitioners expressed significant frustration with the tools already available to them – indeed, their simplicity and rudimentary-ness in some ways played in their favour: practitioners seemed content that audiences understand the visual language of the blurry oval and it is not characterised as ‘deceitful’ in the same way that deepfakes are. Deepfakes – regardless of the intention with which they are utilised – were still viewed strongly by some practitioners as disruptors of the core principles of journalistic work. As a result, the majority of practitioners interviewed expressed some degree of reticence towards the idea that the technology might become more mainstreamed as a journalistic tool for anonymisation. The study suggests that,

while deepfakes have the capacity to be greatly effective in certain circumstances, caution should be taken in not overstating their ‘revolutionary’ capacity to the industry as a whole.

Despite acknowledgement that documentary film and video journalism entails some levels of manipulation – such as the choices of in-camera blurring, framing, or heavy backlighting, as articulated by Sobchack (2004) in her assertion of centrality of the blending of the real and the unreal in cinematic culture – practitioners expressed preference for employing as minimal manipulation as possible in order to tell the stories. Thus, despite Paris and Donovan’s (2019) assertion that the ‘truth’ of audiovisual content has never been stable, this study demonstrates that practitioners still conceptualise truth and realism as fundamental to their work.

This study has been a preliminary look at some of the benefits and limitations: due to the small sample size, it is naturally highly limited in its generalisability – the in-depth perspectives of eight practitioners can by no means speak to the entire industries of journalism or documentary filmmaking as a whole. Moreover, the limited precedent of the use of deepfakes in this context creates a smaller reference point – should they proliferate more, then perhaps in five years' time the results of this study would look very different, as much of the data provided by the practitioners hinged on the novelty of the use of synthetic media in the journalist’s toolkit. What this study has sought to do, however, is to identify some of the contours of debate around the use of deepfakes for this purpose, in order to provide hypotheses for expansion and qualification through future research. As such, some potentially fruitful lines of enquiry might be: How does age and experience impact journalists’ willingness to integrate AI-assisted tools into their practice? How do considerations in anonymisation practices differ in media production in the Global North in comparison to the Global South? How do newsrooms, documentary filmmakers, and freelance video journalists conceptualise risk to sources and their correspondent responsibilities? How do audiences perceive hybrid modalities in journalistic storytelling, and what factors serve to increase or decrease their tolerance for blended formats? What can the formats of docudrama or animated documentary reveal about the intersection between manipulation, creativity and realism within the genre-based norms of non-fiction storytelling?

Conclusion

Welcome to Chechnya's use of synthetic images to anonymise sources experimented with a new tool in the catalogue of post-production methods for the purposes of creating a disguise. By doing so, it subverted an array of normative understandings of the meaning of the image, and the practice of documentary filmmaking and video journalism: it consciously and deliberately uncoupled the image from its referent, for the purpose of the security and safeguarding of the source depicted. It used a journalistic nemesis – the deepfake – to tell a journalistic story, giving it a rhetorical status in and of itself: the use of the deepfake, or digital mask, served to underscore both the material danger that the sources were in, and preserve the facial movements and emotion of their testimony.

This study sought to explore practitioner perceptions of the use of synthetic images – deepfakes – as a tool for anonymisation. In gathering qualitative data on video journalists' and documentary filmmakers' preliminary assessments of the benefits and limitations of the use of AI-assisted technology in this context, it aimed to gain a more practical understanding of the circumstances in which this technology might be utilised in the future, as well as open up future avenues for research on synthetic media's place *within* journalistic production, and how hybrid modalities and advanced creative tools challenge and subvert realist norms of documentary-making.

Through thematic analysis, the study identified four themes through which assessments on the benefits and limitations of the use of deepfakes as tools for anonymisation were made by practitioners: their responsibilities and relationships with their sources, practical and resource considerations, the aesthetic impacts of the image, and wider industry implications. The limitations were more dominant in the data than the benefits: interviews revealed that excessive manipulation was largely problematised, particularly when framed within the context of trust with audiences and the journalists' normative role as 'truth-teller'. As such, while Welcome to Chechnya was widely celebrated as a 'game-changer' in critical circles, this study indicates that a degree of resistance to the uptake of deepfakes as a tool for the purposes of anonymisation of sources remains.

Bibliography

- Adjer, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats and Impact* (pp. 1–27). Deeptrace Labs.
https://regmedia.co.uk/2019/10/08/deepfake_report.pdf
- Allen, W. (1986). Legal tests of photography-as-art: Sarony and others. *History of Photography*, 10(3), 221–228. <https://doi.org/10.1080/03087298.1986.10443113>
- Arrese, Á. (2021). “It’s Anonymous. It’s *The Economist*”. The Journalistic and Business Value of Anonymity. *Journalism Practice*, 15(4), 471–488.
<https://doi.org/10.1080/17512786.2020.1735489>
- Associated Press. (n.d.). *Anonymous Sources*. Retrieved 20 July 2022, from
<https://www.ap.org/about/news-values-and-principles/telling-the-story/anonymous-sources>
- Bakker, P., Broertjes, P., van Liempt, A., Prinzing, M., & Smit, G. (2013). “THIS IS NOT WHAT WE AGREED”: Negotiating interview conditions in Germany and the Netherlands. *Journalism Practice*, 7(4), 396–412. <https://doi.org/10.1080/17512786.2013.802475>
- Barthes, R. (1977). *Image-Music-Text* (S. Heath, Trans.). Farrar, Strauss and Giroux.
- Barthes, R. (1982). *Camera Lucida: Reflections on Photography* (R. Howard, Trans.). Hill and Wang.
- Bartlett, W. (Director). (2020, December 25). *Deepfake Queen: Alternative Christmas Message*. Channel 4. <https://www.channel4.com/programmes/alternative-christmas-message>
- Bate, D. (2013). The digital condition of photography: Cameras, computers and display. In M. Lister (Ed.), *The Photographic Image in Digital Culture* (pp. 77–95). Routledge.
- Bazin, A. (1960). The Ontology of the Photographic Image (H. Gray, Trans.). *Film Quarterly*, 13(4), 4–9. <https://www.jstor.org/stable/1210183>
- BBC News. (2021, January 25). Channel 4 Deepfake Queen complaints dropped by Ofcom. *BBC*.
<https://www.bbc.com/news/technology-55799653>
- Berger, J. (Ed.). (1973). *Ways of seeing*. Viking Press.
- Blackall, M. (2020, December 24). Channel 4 under fire for deepfake Queen’s Christmas message. *The Guardian*. <https://www.theguardian.com/technology/2020/dec/24/channel-4-under-fire-for-deepfake-queen-christmas-message>
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage Publications.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Campbell, D. (2014). *The integrity of the image: Current practices and accepted standards relating to the manipulation of still images in photojournalism and documentary photography*. World Press Photo Academy. https://www.worldpressphoto.org/getmedia/43578992-5421-4e4c-a826-3b993b0b3f3c/Integrity-of-the-Image_2014-Campbell-report.pdf
- Chapman, T. (2022, February 5). The Book of Boba Fett fixed a big Mandalorian problem. *Radio Times*. <https://www.radiotimes.com/tv/sci-fi/book-of-boba-fett-deepfake-comment/>
- Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6), 4–8.
- Chipman, J., Ferraro, M., & Preston, S. (2019, December 24). First Federal Legislation on Deepfakes Signed Into Law. *JDSupra*. <https://www.jdsupra.com/legalnews/first-federal-legislation-on-deepfakes-42346/>

- Cole, S. (2017, December 11). *AI-Assisted Fake Porn Is Here and We're All Fucked*.
<https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn>
- Cole, S. (2018, January 24). We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now. *Vice*. <https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley>
- Cover, R. (2022). Deepfake culture: The emergence of audio-video deception as an object of social anxiety and regulation. *Continuum*, 36(4), 609–621.
<https://doi.org/10.1080/10304312.2022.2084039>
- de Ruiter, A. (2021). The Distinct Wrong of Deepfakes. *Philosophy & Technology*, 34(4), 1311–1332.
<https://doi.org/10.1007/s13347-021-00459-2>
- Deschin, J. (1960). Photography as Art. *Art Education*, 13(6), 7–10.
<https://doi.org/10.1080/00043125.1960.11653780>
- Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, 23(7), 2072–2098.
<https://doi.org/10.1177/1461444820925811>
- Doyle, C. (2011). Documentary. In *A Dictionary of Marketing (3 ed.)*. Oxford University Press.
<https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095724407>
- Duffy, M. J., & Freeman, C. P. (2011). Unnamed Sources: A Utilitarian Exploration of their Justification and Guidelines for Limited Use. *Journal of Mass Media Ethics*, 26(4), 297–315.
<https://doi.org/10.1080/08900523.2011.606006>
- European Commission. (2021, April 21). *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL: LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>
- Fernandez, A. (2022, March 23). Regulating Deep Fakes in the Proposed AI Act. *Media Laws EU*.
<https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/>
- Flick, U. (2018). Doing Qualitative Data Collection – Charting the Routes. In *The SAGE Handbook of Qualitative Data Collection* (pp. 3–16). SAGE Publications Ltd.
- France, D. (Director). (2020). *Welcome to Chechnya*. HBO.
- Frosh, P., & Pinchevski, A. (2011). *Media witnessing: Testimony in the age of mass communication*. Palgrave Macmillan.
- Gershon, I., & Malitsky, J. (2010). Actor-network theory and documentary studies. *Studies in Documentary Film*, 4(1), 65–78. https://doi.org/10.1386/sdf.4.1.65_1
- Given, L. M. (Ed.). (2008). *The Sage encyclopedia of qualitative research methods*. Sage Publications.
- Gladney, G. A., Shapiro, I., & Ray, R. (2013). Reasons for Veiled Sources Spike after 2004 Scandals. *Newspaper Research Journal*, 34(2), 36–49. <https://doi.org/10.1177/073953291303400204>
- Gosse, C., & Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, 37(5), 497–511.
<https://doi.org/10.1080/15295036.2020.1832697>
- Grady, J. (2004). Working With Visible Evidence: An Invitation and Some Practical Advice. In C. Knowles & P. Sweetman, *Picturing the Social Landscape: Visual Methods and the Sociological Imagination* (1st ed., pp. 18–31). Routledge.
- Gray, L., Wong-Wylie, G., Rempel, G., & Cook, K. (2020). Expanding Qualitative Research Interviewing Strategies: Zoom Video Communications. *The Qualitative Report*, 1292–1301.
<https://doi.org/10.46743/2160-3715/2020.4212>

- Grobar, M. (2021, March 2). How VFX Supervisor Ryan Laney Developed Game-Changing Technology For ‘Welcome to Chechnya,’ Shielding Identities Of LGBT Refugees. *Deadline*. <https://deadline.com/2021/03/welcome-to-chechnya-vfx-supervisor-ryan-laney-hbo-documentary-interview-news-1234701188/>
- Hammersley, M., & Atkinson, P. (2019). *Ethnography: Principles in practice* (4 Edition). Routledge.
- Heilweil, R. (2020, June 29). How deepfakes could actually do some good. *Vox*. <https://www.vox.com/recode/2020/6/29/21303588/deepfakes-anonymous-artificial-intelligence-welcome-to-chechnya>
- Jackson, E. (2020, December 10). Face swap: France’s top soap uses ‘deepfake’ technology for self-isolating actress. *France24*. <https://www.france24.com/en/tv-shows/encore/20201210-face-swap-france-s-top-soap-uses-deepfake-technology-for-self-isolating-actress>
- Jones, S. (2017). Mediated immediacy: Constructing authentic testimony in audio-visual media. *Rethinking History*, 21(2), 135–153. <https://doi.org/10.1080/13642529.2017.1305726>
- Kapur, A., & Ansari, N. S. (2022). Coding reality: Implications of AI for documentary media. *Studies in Documentary Film*, 16(2), 174–185. <https://doi.org/10.1080/17503280.2022.2048235>
- Kimball, M. B. (2011). Granting Sources Anonymity Requires Complex Process. *Newspaper Research Journal*, 32(2), 36–49. <https://doi.org/10.1177/073953291103200204>
- Lassiter, L. E. (2005). *The Chicago guide to collaborative ethnography*. University of Chicago Press.
- Lister, M. (Ed.). (1995). *The photographic image in digital culture*. Routledge.
- Lister, M. (Ed.). (2013). *The photographic image in digital culture* (Second edition). Routledge, Taylor & Francis Group.
- Lucas, K. T. (2022). Deepfakes and Domestic Violence: Perpetrating Intimate Partner Abuse Using Video Technology. *Victims & Offenders*, 17(5), 647–659. <https://doi.org/10.1080/15564886.2022.2036656>
- Minichiello, V. (Ed.). (1992). *In-depth interviewing: Researching people* (Reprint). Longman Cheshire.
- Mitchell, A., Gottfried, J., Stocking, G., Walker, M., & Fedeli, S. (2019). *Many Americans Say Made Up News Is a Critical Problem That Needs To Be Fixed* (p. 74). Pew Research Center.
- Mitchell, W. J. (1994). *The reconfigured eye: Visual truth in the post-photographic era*. MIT Press.
- Mitchell, W. J. T. (1986). *Iconology: Image, text, ideology*. University of Chicago Press.
- Mortensen, T. M. (2021). *The Ethics of Photojournalism in the Digital Age*: By Miguel F. Santos Silva and Scott A. Eldridge II, Milton Park, UK, Routledge, 2020, 162 pp., \$41.20 (pbk), ISBN-13: 978-1138586307. *Digital Journalism*, 9(1), 126–127. <https://doi.org/10.1080/21670811.2020.1791205>
- Murray, M. (2018). Narrative Data. In *The SAGE Handbook of Qualitative Data Collection* (pp. 264–279). SAGE Publications Ltd.
- Newton, O. B., & Stanfill, M. (2020). My NSFW video has partial occlusion: Deepfakes and the technological production of non-consensual pornography. *Porn Studies*, 7(4), 398–414. <https://doi.org/10.1080/23268743.2019.1675091>
- NVIDIA. (2022). *Random Face Generator (This Person Does Not Exist)*. <https://this-person-does-not-exist.com/en>
- Panetta, F., & Burgund, H. (2019). About The Project. *In Event Of Moon Disaster*. <https://moondisaster.org/about>
- Paris, B., & Donovan, J. (2019). *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*. Data & Society. https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf

- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (Fourth edition). SAGE Publications, Inc.
- Peirce, C. S. (1985). Logic as Semiotic: The Theory of Signs. In R. E. Innis (Ed.), *Semiotics: An Introductory Anthology* (pp. 4–22). Indiana University Press.
- Peters, C., & Broersma, M. J. (Eds.). (2013). *Rethinking journalism: Trust and participation in a transformed news landscape*. Routledge.
- Pink, S. (2013). *Doing visual ethnography*.
- Pjesivac, I., & Rui, R. (2014). Anonymous sources hurt credibility of news stories across cultures: A comparative experiment in America and China. *International Communication Gazette*, 76(8), 641–660. <https://doi.org/10.1177/1748048514548534>
- Popova, M. (2020). Reading out of context: Pornographic deepfakes, celebrity and intimacy. *Porn Studies*, 7(4), 367–381. <https://doi.org/10.1080/23268743.2019.1675090>
- Postema, S., & Deuze, M. (2020). Artistic Journalism: Confluence in Forms, Values and Practices. *Journalism Studies*, 21(10), 1305–1322. <https://doi.org/10.1080/1461670X.2020.1745666>
- Ritchie, J., & Lewis, J. (Eds.). (2003). *Qualitative research practice: A guide for social science students and researchers*. Sage Publications.
- Ritchin, F. (2009). *After photography* (1st ed). W.W. Norton.
- Rosner, H. (2021, July 17). The Ethics of a Deepfake Anthony Bourdain Voice. *New Yorker*. <https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice>
- Roulston, K., & Choi, M. (2018). Qualitative Interviews. In U. Flick, *The SAGE Handbook of Qualitative Data Collection* (pp. 233–249). SAGE Publications Ltd. <https://doi.org/10.4135/9781526416070.n15>
- Sample, I. (2020, January 13). What are deepfakes – and how can you spot them? *The Guardian*. <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>
- Santana, A. D. (2014). Virtuous or Vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, 8(1), 18–33. <https://doi.org/10.1080/17512786.2013.813194>
- Schleser, M. (2022). Smart storytelling. *Studies in Documentary Film*, 16(2), 97–113. <https://doi.org/10.1080/17503280.2022.2048230>
- Schreier, M. (2018). Sampling and Generalization. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Collection* (pp. 84–97). SAGE Publications Ltd.
- Scott, C. R., & Rains, S. A. (2020). (Dis)connections in anonymous communication theory: Exploring conceptualizations of anonymity in communication research. *Annals of the International Communication Association*, 44(4), 385–400. <https://doi.org/10.1080/23808985.2020.1843367>
- Shin, S. Y., & Lee, J. (2022). The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes. *Digital Journalism*, 10(3), 412–432. <https://doi.org/10.1080/21670811.2022.2026797>
- Snyder, J., & Allen, N. W. (1975). Photography, Vision, and Representation. *Critical Inquiry*, 2(1), 143–169. <https://www.jstor.org/stable/1342806>
- Sobchack, V. C. (2004). *Carnal thoughts: Embodiment and moving image culture*. University of California Press.
- Sontag, S. (1978). *On Photography*. Farrar, Strauss and Giroux.

- Tagg, J. (1993). *The burden of representation: Essays on photographs and histories*. University of Minnesota Press.
- Teus Media. (2022). *TeusMedia—Home*. <https://www.teus.media/>
- Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Niessner, M. (2016). Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2387–2395. <https://doi.org/10.1109/CVPR.2016.262>
- Thomson, P. (2020, June 30). Digital Disguise: ‘Welcome to Chechnya’'s Face Veil Is a Game Changer in Identity Protection. *Documentary Magazine*. <https://www.documentary.org/column/digital-disguise-welcome-chechnyas-face-veil-game-changer-identity-protection>
- Usher, S. (2021). *Letters of note: Correspondence deserving of a wider audience*. <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2908472>
- van der Nagel, E. (2020). Verifying images: Deepfakes, control, and consent. *Porn Studies*, 7(4), 424–429. <https://doi.org/10.1080/23268743.2020.1741434>
- Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes. *Journalism Practice*, 15(6), 803–820. <https://doi.org/10.1080/17512786.2021.1908838>
- Wallsten, K., & Tarsi, M. (2016). PERSUASION FROM BELOW?: An experimental assessment of the impact of anonymous comments sections. *Journalism Practice*, 10(8), 1019–1040. <https://doi.org/10.1080/17512786.2015.1102607>
- Wiles, R., Coffey, A., Robinson, J., & Heath, S. (2012). Anonymisation and visual images: Issues of respect, ‘voice’ and protection. *International Journal of Social Research Methodology*, 15(1), 41–53. <https://doi.org/10.1080/13645579.2011.564423>
- Williams, L. (1993). Mirrors without Memories: Truth, History, and the New Documentary. *Film Quarterly*, University of California Press, 46(3), 9–21.
- Winston, B. (2000). *Lies, damn lies and documentaries*. BFI Pub.
- WITNESS (Director). (2020, September 14). *Identity protection with deepfakes: ‘Welcome to Chechnya’ director David France*. <https://www.youtube.com/watch?v=2du6dVL3Nuc>
- Wolfe, J. (2021). Crossing the Line: How the “Roadrunner” Documentary Created an Ethics Firestorm. *NAB Amplify*. <https://amplify.nabshow.com/articles/crossing-the-line-how-the-roadrunner-documentary-created-an-ethics-firestorm/>
- Zylinska, J. (2017). *Nonhuman photography*. The MIT Press.

Appendix

Interview | Ryan Laney

Nathalie Weatherald (NW): Tell me about how you approached the development of the visual effects.

Ryan Laney (RL): Development was out of necessity.

As you know, David France, he read an article originally by in the New Yorker by Masha Gessen. I don't know if you know that name. But, she had kind of reported on this, and it wasn't really well known - anyway, he jumped on a plane and went over and got access to an underground railroad, which is what was covered in Welcome to Chechnya. Yeah. And he shot for a year. And he had a good solid rough cut. And no way to take it to market. Yeah, so that's when, through that entire year that they shot, their first kind of tour over there, they have been trying all sorts of different ways to disguise people.

But David, because of this particular group that he's highlighting the trials of, it's very easy to anonymise the idea of a person and that kind of dehumanises them at the same time. And he wanted to maintain that human connection. So, when he came to us, his original idea was doing something like Scanner Darkly or in more modern terms. What is it, 'undone'? Which is an effect, sort of a cartoon effect over the footage, called rotoscoping.

So, we were introduced through a mutual friend and he came to us and said, can you do it? Can you can you automate this so that, one, we can afford to do it, and two, to not have to send it to a rotoscope farm somewhere, since we don't know who's gonna be seeing it or working on it, and they don't have security protocols in place. So that was my introduction to David and Alex Henty, the producer on the show.

And at that point, we did a test with style transfer, which is a machine learning tool that kind of applies artistic effects to video. We did a test with that. And the first thing that we noticed from their previous hands-on test, and this is that caricatures of people often accentuates the things that make them unique versus hiding them.

And so, we saw right off the bat that we needed to switch gears.

And so, we call, show them what we've done. And we're like, this is why we think this isn't gonna work. Then we have this other idea - if we can kind of move the bones underneath somebody's face, that will change who they are, like, fundamentally, their DNA in a sense, right?

If you have, I mean, think of any movie where you've seen a transformation of a person into a creature, you know, like the werewolf films, you know, something like that. The way that works is you get a 3D track of the head, and then you map the face onto it and you pull the vertices around. And, again, if you use kind of like human anatomy to make those moves, you can do it in a believable way.

That was the first case. And David said, we love it. But we also have to make this so good that even their parents wouldn't recognise it. Because it looked like the cousin of the person.

So, we had to, we had to take it a step further. And we rolled in this idea of style transfer, which is where you apply an artistic effect of somebody onto somebody. And we said, well, what happens if, instead of fitting in an artwork, we fit in another person's face?

And so, what the neural network does is it kind of separates what the face is doing - are the eyebrows up or down, are the eyes open, or closed - from what they look like. So, do they do have long eyelashes or short eyelashes? How are your eyebrows shaped, versus where are they.

So, in that sense, we're able to take those, those things that are very unique, and the concert of facial features in the face or in the organisation they are laid out, is what makes us unique. So, we were able to sort of combine these two ideas of moving the bones under the skin, and then changing the shape of eyebrows and lips to in no shapes to be able to effectively find that solution, which was so their parents wouldn't recognise them.

NW: So, would you call the technology you developed a 'deepfake'?

RL: Yes and no. There was a research paper that informed us that I think also informed deep fakes. So, they are both using neural networks in order to, to do things in the code, the deepfake code that I've looked at – we didn't use those tools for a couple of reasons.

One, code bases were binaries that came out of Russia, and we didn't feel comfortable installing... We don't know what's in it, right? It's not transparent. But if we code our own, we use Python and Google library called TensorFlow. If we code our own, we know what's inside, and we can at least audit what's going on.

So, the bigger thing for fitting against the requirements is deepfakes, they do this thing where they shift the pixels around in order to fit the faces better. What this does, is you get an identification transference, right? You lose that change in shape of the layout. So, when we do it, you end up where the eyes aren't exactly in the same place. A deepfake is more likely to have this identification transference, because it's doing a really good fit, for fidelity of picture, but we actually have more interest in hiding. We don't have any layers that cause that transference across, we can keep things separated in that way.

And then kind of back to the original question is, is this a deepfake? We think there's a kind of a philosophical conversation about it.

So, I've got 30 years in visual effects. If we were to say that a deepfake is anything synthesised with a neural network, then we say that, you know, in a mathematical identity, we say that, if that's true, then anything synthesised is fake.

And that means that everything in every movie I've ever seen, when you - I don't know if it's on Meet or on Zoom, they have the little button that says 'touch up my appearance' that uses removal of backgrounds - to use machine learning to know how to apply those effects, in which case, any video conferencing call is a false representation? Or is it, right?

So - we're both being transmitted through a compression algorithm right now, optimised by some machine learning protocol. So you just need to be careful about where that line is drawn.

So, there was a high-profile court case in the US this year where the prosecution tried to hand the witness an iPad, for purposes of identifying something in an image. The Defence objected and said it's on an iPad and iPads use machine learning and therefore that is not truthful in what it's representing. The judge sustained it. So, there is actually now legal precedents in the US that anything which uses machine learning, such as an iPad, which is presenting a video, is not real.

So then where do you go, right? Then nothing's real.

So, I think it's a really interesting idea in the bigger picture, where do you draw that line? I think it comes down to intent. And when the intent is obvious, showing somebody a picture on an iPad does not intend to be deceiving.

Obviously, what we're doing - every film we've worked on has a disclaimer, there's a halo that we put around the work so that even if you didn't have the disclaimer, you could be able to read and see, the audience can really see that there's something going on.

And we've tied that visual language to the blurry oval to also say that, you know, this is the language of disguising identities.

So, is that fake? Because what we're actually interested in is this fidelity of motion in the face. And we have a very high mark for that. So, we have high marks for fidelity of motion and the transference of emotion, then we actually have a truthful representation of what was originally there, even though the picture is different.

So, David wanted to call them 'deptruths' because of this idea that we're actually allowing the audience to see something that is more truthful than darkened silhouette or a really blurry face, right, where there's no indication of what was actually there. We're giving us sort of this more honest representation than anything else that could be done. So yeah, it's really interesting.

As the conversation evolves, legally speaking, I think there's a lot of, of laws already on the books around fraud and deception.

NW: Have you been hit by any of them? Have any of them touched your work?

RL: No, no, we had a really warm welcome. We haven't had a single - that I'm aware of - a single negative response to it. In fact, one of the film's funders, Jesse Ferguson, was on Jimmy Kimmel. For his, I think, a book he wrote or another project he's working on. And he mentioned Welcome to Chechnya. And he uses the term deepfakes and Jimmy Kimmel's response was "it's the one good use."

So, there was this project, called In Event of Moon Disaster, I don't know if you've seen it, but he does this a really interesting thing where he presents this video, and it's on the topic of deepfakes. And he shows you some they recreated a speech by Nixon giving us the speech that was actually written - like it was a real speech - that was written in case something went wrong with the moon landing.

So, they recreated this sort of like, alternate history. It's the same setting is that he gave the moon landing presentation, except it was a different stage, this other speech. So, they do this, and they ask you a few questions. What did you see that's changed, etc, and then they show you all the things.

So, with deepfakes the technology that made them famous was used for non-consensual porn, and which in some places is considered sexual assault, and in some cases, the law doesn't care at all. There's some sort of high-profile moves in the UK right now to try to get legislation. I read about it.

I don't understand how if somebody goes on television, and he says something is not true about somebody they can be sued for slander, but, in most places, if somebody goes on television, and shows a picture of somebody doing something that they didn't actually do - I don't see a difference. I don't see where the current laws don't hold up to this new technology. And maybe people just need to say like, oh, well, visual language is visual language, and therefore all laws that applies as language should apply to visual language.

So, I feel like there's a lot of people that are like, well, we got to outlaw, we got to ban, and we got to do whatever. And I think that there's a lot of laws that should already cover this, if they were looked at from this lens of 'video is language which is communication, and therefore laws apply and communication should apply to a video'. So yeah, it's really interesting.

There was this really famous talk on I think, Crossfire, it was a PBS television show in the 80s or 90s. And Frank Zappa got on the show, surrounded by really conservative guys to talk about, what at the time was a hot topic of banning songs that use certain words? Okay. And Frank Zappa must have said 100 times they're just words.

It's not the words that people are afraid of. It's the context. And I think that if we can, if we can move the conversation away from you know, what is this technology to what is this? What is this implementation of this technology doing? Then I think we get closer to an answer of how to legislate or how to have those conversations.

But as long as we say, you know, this thing is bad. Yeah. Something else will just pop up. That's just legally different enough to not be that thing, but more terrible because it had to evolve to be. Yeah, to get around the cracks.

NW: Have you had any roadblocks in the development?

RL: That part of the road went totally smoothly, like once we saw it work. I remember getting chills the first time. The first time it worked. I was on the phone with David and Alice. And he's like, Oh, my God, this is going to change witness filming. So, we did have a good idea right away.

However, we were concerned about one thing that was this uncanny valley, that's often a conversation topic in visual effects and how close can you get to human without, you know, if you're not if you're not close enough? It's off putting.

So, David hired Talia Wheatley in a Dartmouth College and she did a she did a study. And through that we had some high level of confidence also that this was the right we tried several things through her. And we felt like this was the right solution for that. As far as being enough. David was in dialogue with some of the contributors to the film and they felt like it was enough. One contributor, said when he saw the film, he felt like he was watching somebody else go through it. And that helped him process the trauma better.

We showed to Abraham - the first person we tested with. He's a friend in Los Angeles and said, Hey, we got this crazy idea, would you come sit for us? So, we did this little test. And we showed him and he's like, Oh, that's, that's really strange, because he's kind of my build. And I know that, but I know that's not me. So, he had a similar. Like, even though he knew it wasn't him felt like he was watching himself. So, with those kinds of perspectives of people who were very close to material, we felt like we did a good job of masking.

It turns out, Maxim, who's now known to the world, through the film, Maxim and Abraham are about a foot difference in height. But they're the same kind of build, like within their height, so it worked out well.

We didn't know what to expect [when we released the film], but certainly eager to see how people react. We went to Sundance, and there was a screening there. And in the Sundance screening, I had a couple of test clips that were related to the film on my phone, and just walked around. And you know, why are you here? Or we're working on this project? And can I get your opinion on this?

And all of that was like, Oh, wow, this is this is very different. So when we went and saw the film, there wasn't as much sort of concern about it, because we had these, you know, several days leading up to the first showing of the film. Yeah, it kind of trialled in town and talking to Nikki sort of getting feedback, because we were I mean, we had to lock down this, I mean, this was before the pandemic, but we had to lock down because of security. So, we were kind of like, in a bubble already.

So one interesting thing in the in the first screening, you know, so there's a disclaimer, and then the first 20 shots, we did this thing where we played out the softness, and also, the Assistant Editor, Maxwell Anderson, had gone in and taken dialogue out of the first shots, so the people weren't reading subtitles. And they could read the disclaimer, and then see faces and focus, and understand the interesting effects and softness up close. And so that we had a lot of back and forth a lot in how we acclimate the audience so that they understand and this is a big part of, again, like the conversation of what is real and what is fake is having the dialogue with all parties, participants, David was in dialogue with the participants and we're in dialogue with the audience. And so, everything was kind of open.

So, there didn't seem to be any concern about the acclimation once people were in the film. You very readily just watch it as an audience member. You forget about you forget it there. Yeah. And then David, obviously, you've seen it. So, in this very specific moment, the press conference to be held comes off. And the audience gasped. It was so weird when I watched.

NW: Yeah, because you get so used to a character and I almost then it took me a minute to emotionally connect with this new face. Because for the duration of the film, I've been emotionally connecting with this other face, and to kind of shift the character who you're connecting with.

RL: Yeah, it's, it's funny that you say that because that was David's number one concern, really for a while. There was a discussion about Maxim being without a veil and all the people around him would be with veils, and he felt like that was not really authentic to the moment it was filmed because before the press conference, he was not known and after the press conference, he was known so, that was the, that was the time to do it. But he was concerned that he's like, Abraham is so cute and people are gonna love him. And then they're not gonna know what to do with this new person. But we decided that Maxim was cute also, and so it was gonna be okay.

But I think the emotive power of that, because it's kind of meta in a way because you're you also then become aware that you've been watching a disguise, which has been necessary for this duration. So, there's a kind of extra level that you're playing with.

Yeah, I think it actually did play into kind of the reception of it. Because without that people might have said, no, I'm not realising there's anything going on.

So totally normal, just a bit fuzzy.

NW: I read that you're working on new projects. Can you tell me what they're about?

China and North Korea, women's rights, human trafficking, dark money. Wrongful imprisonment.

We've consulted on about 50 or 60 projects at this point? Well, we haven't we haven't done all of them. But we've had conversations about how, you know, how to go about doing the work or how we are limited in our bandwidth. So, we've also shared a lot of what we do with people to help them to their own projects.

So yeah, we and we're currently doing talks and whatnot. We did a workshop with MIT open doc labs, right after the show. And the sort of the overwhelming consensus was, one, you got to put some gates on this.

And that it needed to be something that was available, not just to the biggest projects, but also to basically anybody that needed it. So, our goal, our sort of, like, pressure has been on how to automate things even more, like to be able to do it for greater numbers.

It was more shots than were in the original Matrix.

The production management of it was as effortful as, you know, a film of that scale. It's 400 shots. So, 400 shots is a small project. And the typical shots are like, you know, a Marvel film is going to have 2000 shots, which is basically every shot in the film.

But to do 400 shots on a documentary, missing two zeros in the budget. Yeah, you know, so we're already kind of in that, in that ballpark.

Anyway, we've put a lot of work into the production to so that they can shoot some of the data and they have more knowledge and how that stuff goes, they can do some of the prep work and that's helping bring that number down, also make it again, more available to films that otherwise wouldn't, wouldn't have access. So our goal is to make sure that you know, it gets out there but it's not just free willy nilly.

NW: Has the technique developed or evolved at all since Welcome to Chechnya, through the work on these new projects?

We've improved the resolution, the sort of the pixel quality.

There's a couple of things that we sort of learned on Welcome to Chechnya that we were able to like oh if we just did if we you know step this way a little bit, we can get a better result here. If we add this piece of information, we're getting better result there.

One thing for instance, in Welcome to Chechnya we shot nine cameras - we didn't know where it was gonna break. So, we overlay over again, everything. We shot nine cameras running, there was an issue with the camera triggers. So, they couldn't be remotely triggered. They hadn't actually had to go up and click. For the doubles.

And so, we've simplified that process. And we don't actually need that many cameras.

We didn't know what we didn't like what was going to work and what wasn't yet. So, when initially we probably shot for maybe an hour or two per person, we can now shoot in about 15 minutes per person. So, we've gotten we've gotten the sort of like, requirements of, again, by learning what worked and what we need, and whatever. The thing that in in that first round of data capture. You know, we're figuring out there, they're not actors. So, a lot of they're there while they're sitting there looking off in one direction where the director is. And so, like our whole data set was biased towards this one point of view.

So, if we just if we had just, you know, put all the data in as it came in, gotten the bias in certain eyeline that is above centre and a certain view that is, you know, off to one side. So, so in the in the sense of like, how will we improve some things with like, oh, well, if we tell the machine which way the eyes are looking, we can pair between the film and the data, and make sure we have the right eyes, so we get a clearer version of the eyes. So, a lot of little just technical, technical things like that.

Again, like where we can split labour between what we're doing and what the clients doing. They can use their resources and reduced overall costs.

We didn't go the path where it will be patented. We don't hold any rights or copyrights over the finished material, it's a normal project where work for hire, we provide a service.

idea of patenting is an interesting one. But by the time we finished the project, like the industry had moved forward, the year after Welcome to Chechnya came out, Google presented a paper that was pretty identical to what we were doing.

And also, we had some risk if we tell people exactly what we're doing. Because what we're doing is we're creating an encryption key.

We're encrypting one phase, and we're doing that in a very secretive... like it's better to be a secret sauce than it is to be a patent.

You know, our market is a documentary film. It's not like we've invented something for the oil industry.

We've got a small market and they don't have a budget.

And it was also countered to that first conversation that we had with MIT where this needs to be available for everybody the way you make it available for everybody is you talk about how you did it, and they can build their own. The tools exist. TensorFlow is open source, Python is open source, like the tools we're using exist in the world. And anybody with a visual FX background who has the inclination, could actually do its work. And they're starting to they're starting to do more like, in the range of face replacements, Lucasfilm did something that maybe wasn't as successful and in the Book of Boba Fett, it was improved on, so we'll see, I think people are understanding it's, it's a hard thing to do.

But we do encourage other people to do it. Because I feel like that the more people there are doing it, the more that drives the cost down in the more not necessarily competitive, like you have to free but in in the oh, I'm doing it, I figured out this thing. And it makes it that much more efficient. And therefore, yeah. So by furthering the technology, and the things we can, we can work on cost.

And also, the more people who are doing it for this market, the more of the market can be served, because we are we are a limited number of brains. Yeah.

I think I saw something that said 60 to 70% of all content on the internet has video. I don't know what by what metric if they were, you know, megabytes or something, or reading time versus watching time. But they're projecting that is, is on an increasing tangent. So, the medium will be continued to be more original, which, if we think about it. We're humans, and we're tribal animals. Yeah, we, we want to have visual communication. We're not really engineered for reading, we're engineered for communicating in person. And so if, I mean, video is kind of a bad proxy, but it's maybe there's a group people that will learn better through video than through reading? I don't know, it's hard to say.

When we thought about the implications in that first call with David, we asked: well, who invented the blurry oval?

It was just suddenly used everywhere. So, I mean, we didn't really invent everything we just applied style transfer to in a very, very specific domain. So yeah, we're happy to have contributed something

to the world, contribute something valuable to the world. And I feel like, my work is not done. I gotta, now that I have a taste of that. Got to do more.

I mean, my interest, lies in philosophy. And so, I was very excited about just the conversation about, you know, what is right in this and what is on the edge of this, so, I love those sorts of pedantic conversations about really nothing. For me, it's been having a taste of doing something that's valuable in the world. It's not to say that summer blockbusters aren't valuable, but it's a different, it's a different conversation.

I've already worked for ILM, and Sony in digital domain. So, I've already worked with the best people in the industries. But if one of those companies were to come back and say, come back, I would have to think about it whereas before it would be absolutely yes. Because they're great companies and they do they do the best work that exists in you know, creating alternate realities and I just I feel like there's such a desire to kind of leverage this. Like, oh, we can we can actually do a lot more here. We can add a lot more value in kind of a real way.

So, yeah, I think it's changed me and if I can get through the rest of my life and never do fictional work, probably that will be fine with me too.

I'm having these conversations in these places where they're not allowed to have encrypted drives, and they don't have power 24 hours a day, and in order to get a message out, they gotta go through a coyote who's going to take all their money and not actually pass the message on, and, and you start to look at like - wow, there's a lot of barriers, not just like going to market, there's a lot of barriers to getting information into and out of these places.

And so, for me, I've been starting to have some conversations about like, how might you sort of solve the problems in the enabling storytellers, as much as you are enabling them to tell their story like, so right now.

NW: Thank you for your time!

RL: You're welcome.