ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/139988/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Wahl-Jorgensen, Karin and Carlson, Matt 2021. Conjecturing fearful futures: Journalistic discourses on deepfakes. Journalism Practice 15 (6) , pp. 803-820. 10.1080/17512786.2021.1908838

Publishers page: http://dx.doi.org/10.1080/17512786.2021.1908838

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Conjecturing fearful futures: Journalistic discourses on deepfakes

(Accepted for publication in *Journalism Practice,* Future of Journalism special issue, March 19, 2021)

Karin Wahl-Jorgensen
Cardiff University
(corresponding author)
Address:
Cardiff School of Journalism, Media and Culture, Cardiff University
Two Central Square, Rm. 2.62
Central Square, CARDIFF, CF10 1FS, Wales, UK
E-mail: wahl-jorgensenk@cardiff.ac.uk

Matt Carlson
University of Minnesota
E-mail: carlson1@umn.edu

Conjecturing fearful futures: Journalistic discourses on deepfakes

Abstract

This paper investigates journalistic discourses on "deepfakes" as the future of fake news. We analyze journalistic discourses on deepfakes over an 18-month period from January 1, 2018, when a GoogleTrends search demonstrates that the term first began to circulate, to July 1, 2019, shortly after an altered video of Nancy Pelosi, the US Speaker of the House, circulated on social media. Based on a comprehensive thematic analysis of English-language news stories on the topic, drawn from Nexis UK, we suggest that journalistic responses to deepfakes reveal deeper anxieties both about the future of the information environment and journalism's role within this environment. Concerns that the audiovisual nature of deepfakes makes them inherently more believable than previous fake news forms leads to worries over the impending weaponization of deepfakes by resource-rich "bad actors." At stake is the trustworthiness of media content, and journalists' role in providing verified content to the public. We argue that journalists conjure up speculative worst-case scenarios around deepfakes – what we refer to as "conjectured specifity" to highlight the vital importance of journalism as a bulwark against fabrication and a defender of truth.

Keywords:

Artificial intelligence, deepfakes, fake news, journalistic authority, misinformation, trust in journalism

Conjecturing fearful futures: Journalistic discourses on deepfakes

Introduction

This paper investigates journalistic discourses on "deepfakes" as the future of fake news. The emergence of deepfakes results from recent technological developments in machine learning in which a program combines two distinct sets of images to create a single fabricated audiovisual image that is difficult to distinguish from unaltered video sources.[1] Stories about deepfakes started to circulate at the beginning of 2018, following on from the popularization of an app that made it easy to fabricate celebrity porn videos. Soon thereafter, the technology was used to manufacture fake videos of politicians and other public figures. In an early high-profile example, director Jordan Peele used the technology to create a video of Barack Obama calling Donald Trump a "dipshit" to warn the public about the dangers of deepfakes (e.g. McCarthy, 2019). In a similar warning about the evolving technology during the run-up to the 2020 US Presidential Elections, the nonpartisan advocacy group RepresentUs released deepfake ads featuring Russian president Vladimir Putin and North Korean leader Kim Jong-Un proclaiming the impending collapse of American democracy (e.g. Hao & Heaven, 2020). Political and social institutions, from the US House of Representatives Intelligence Committee to the European Commission, have voiced concern about the potential harms of deepfakes, ranging from fake declarations of war to sexual harassment and incrimination through fabricated porn videos. Fears of deepfakes are shared by members of the public: A survey by the Pew Research Center found that 63% of US adults expressed concern that "altered videos and images create a great deal of confusion about the facts of current events"

(Gottfried, 2019). Moreover, a majority of respondents supported efforts by social media companies to suppress these videos.

Along those lines, social concerns about the impact of fabrication and AI connect to long-standing preoccupations with the quality of information in the social media environment. Social media platforms radically extend the ability to post and circulate information, thereby dampening the gatekeeping power of traditional journalism (Nielsen & Ganter, 2018). Following the 2016 US presidential election, these concerns coalesced in an informational moral panic around what is widely described as "fake news" – wholly fabricated news reports (Carlson, 2018a). These fabrications – and traditional journalism's loss of control over mediated communication that they symbolize – are central to more foundational debates over journalistic authority and judgment, as well as trust in news (e.g. Carlson, 2009; 2015; 2018b).

Deepfakes enter the media landscape at a time when journalistic authority is already challenged by increased criticism and an environment marked by an expanded range of platforms and public communicators (Vos & Thomas, 2018). Experimental research on deepfakes connects them to greater uncertainty around media content (Vaccari & Chadwick, 2020). Yet this remains a nascent area of inquiry, with research on the social and political implications of deepfakes only starting to emerge (e.g. Dobber et al., 2020; Diakopolous & Johnson, 2019). In this paper, we contribute to the development of such research. We investigate journalistic discourses on deepfakes, arguing that they capture complexities around the potential impact of the new technology to either mislead or increase informational incredulity, while highlighting how journalists publicly support their societal role of policing truth to maintain the legitimacy and authority of journalism.

Notably, deepfake technology remains in the earliest stages of development and use, such that journalists focus on its potentialities for future disruption more than on existing examples or their effects. This news discourse takes on a particular temporality that Neiger (2007) calls "conjectured futures" – prognosticating with an emphasis on instilling fear around worst-case scenarios. Debates over deepfakes – like the broader discussion of fake news – cultivate and mobilize anxieties over the role of journalism in a new and swiftly changing media ecology, where the privileged role of established institutions can no longer be taken for granted. Central to our argument is the idea that coverage of deepfakes differs from previous considerations of fake news due to the primary discursive focus on constructing the technology as a *future threat* rather than a *realized problem*. Stories are rarely based on incidents that have already occurred, in contrast to the focus on actual fake news stories (Tandoc et al., 2019). Instead, journalists construct a vision of the future that then shapes the public understanding of what is to come. In the case of deepfakes, the conjectured future relies on the construction of very specific, yet fictional, worst-case scenarios, or what we refer to as "conjectured specificity." This conjectured specificity, in turn, informs imagined solutions, central to which is the vital role of journalism as an institution.

We analyze journalistic discourses on deepfakes over an 18-month period from January 1, 2018, when a GoogleTrends search demonstrates that the term first began to circulate, to July 1, 2019, shortly after an altered video of Nancy Pelosi, the US Speaker of the House, circulated on social media. Based on a comprehensive thematic analysis of English-language news stories on the topic, we suggest that journalistic responses to deepfakes reveal deeper anxieties both about the future of the information environment and journalism's role within this environment. Concerns that the audiovisual nature of deepfakes makes them inherently more believable than previous fake news forms leads to anxieties over

the impending weaponization of deepfakes by resource-rich "bad actors." At stake is the trustworthiness of media content, and journalists' role in providing verified content to the public.

Contextualizing deepfakes

As a starting point, concerns around deepfake technology should be placed within the historical context of how new technologies bring with them new questions about the veracity of media representations. Incidents ranging from the *New York Sun's* Great Moon Hoax of 1835 (Thornton, 2000) to Orson Welles's 1938 *War of the World*s radio broadcast (Schwartz, 2015) point to unease about the power of fake media content to affect people's beliefs and behaviors. Past fears of fake news caused journalists to reflect on how their practices could be marshaled against fabrications (Creech & Roessner, 2018; Finneman & Thomas, 2018). Recalling this longer history is not meant to dismiss concerns about emerging technologies like deepfakes, but instead to recognize how such concerns manifest themselves as moments of reckoning for journalists about the boundaries between legitimate news and the hoaxes that threaten their authority as purveyors of truthful accounts.

Deepfakes are related to the recent panic surrounding so-called fake news via digital media—wholly fabricated stories mimicking news conventions whose often shocking content is meant to lure engagement for the purposes of profiting from click-based digital advertising and/or affecting public opinion. The panic over fake news first came to the fore in the context of the 2016 US presidential election. A *BuzzFeed* investigation found that "the top-performing fake election news stories on Facebook generated more engagement than the top stories from major news outlets such as the *New York Times*, *Washington Post*, *Huffington Post*, NBC News" (Silverman, 2016). The controversy over the role of fake news also surfaced in discussions around interference in elections on a global basis. Misinformation

campaigns, facilitated through social media, have been alleged in electoral contests ranging from the British EU Referendum in 2016 to the 2018 Brazilian elections which swept right-wing populist Jair Bolsonaro to power (e.g. Grice, 2017; Phillips, 2018).

Just as quickly as the concept of fake news was articulated, observers began questioning the term's usefulness. First, the term has become problematic because of its appropriation by Donald Trump to attack mainstream media for any reporting critical of his administration (e.g. Waisbord, 2018). Second, scholars have argued that the phrase unhelpfully collapses an array of practices that ought to be differentiated according to their intent (e.g. Corner, 2017; Jack, 2017; Wardle & Derakshan, 2017; Tandoc, Lim & Ling, 2018). While some types of misinformation may be shared in error or for satirical purposes, most observers suggest that the kinds of "fake news" we should worry about are those intended to cause harm (Wardle & Derakshan, 2017). The concern is that democratic governance is damaged by the lack of shared trust regarding the veracity of public information. RAND Corporation researchers summarize this danger through the concept of "truth decay," which describes four main trends in the contemporary media environment that have come about due to technological and social transformations (Kavanagh & Rich, 2018). These trends include heightened disagreement about facts and analytical interpretations of data; the blurred line between opinion and fact; the increased volume and influence of opinion and personal experience across the communication landscape, and diminished trust in formerly respected institutions as sources of factual information. Together, these trends highlight the existential threat for conventional news media—and wider threats to the polity—represented by the rise of misinformation.

For journalism scholars, debates over fake news feed into broader anxieties about the changing role of journalism in a swiftly changing media ecology. As Waisbord (2018, p. 1868) argues, fake news is "indicative of the contested position of news and the dynamics of

belief formation in contemporary societies. It is symptomatic of the collapse of the old news order and the chaos of contemporary public communication." Such struggles over truth are ongoing, finding new purchase as contexts change and new epistemological forms arise (p. 1869).

Journalism's professional authority is centrally at stake in these debates. If truth-telling and objectivity are at the heart of conventional understandings of journalism ethics and authority (McNair, 1998; Ward, 2009), the multiplication in opportunities for spreading misinformation presents a distinct challenge to core professional values.

The openness of social media platforms provides journalists with a point of comparison to argue for their value as arbiters of truth (Carlson, 2018b). Social media platforms have greatly expanded the range of mediated voices beyond the restricted channels of mass communication, even as they have maintained a position of being passive distributors of content rather than publishers responsible for their content. However, there is ample debate about what responsibility social media companies should have for the content on their sites and how active they should be in policing this content and those who post it. For example, such questions were prominent in debates over the suspension of Donald Trump's Twitter account in January 2021 (Clayton, 2021). This includes discussion over whether to remove fabricated news stories.

The challenges of identifying and removing fake news on social media – and the platforms' ambivalence about evaluating the accuracy of content – gives journalists an opening to publicly tout their professional verification practices based on their privileged claims to objective truth. In this way, journalists engage in boundary work (Carlson & Lewis, 2015) to support their legitimacy by excoriating social media as unreliable. Journalists thereby "position professional journalism as principled, ethical and fully aware of the role it plays in democracy, and they position Facebook as haphazard, unethical and in denial of the

important role it now plays in democracy" (Johnson & Kelling, 2018, p. 829). Such boundary work is about establishing journalistic authority (Carlson, 2017). Authority is never granted or assumed, but arises through an ongoing relationship of dependence and trust that must be continuously established between journalists and their audiences. The work of establishing journalistic authority arises both through the repetition of news practices and a metadiscourse that engages in interpretive work to define appropriate practices and norms while marking others as deviant or outside of journalism.

Journalism and the Temporality of Threat

In this fraught context, examining how journalists talk about the future information environment, and the threats and opportunities it holds, reveals much about how they defend their authority as trusted information producers. A deeply held assumption concerning journalistic temporality is that news is dominated by accounts of past occurrences, such that the question news stories answer is, "what happened?" Yet much news discourse is actually aimed at the *future*, focusing instead on "'What will happen?' or 'What is likely to happen?' (Neiger, 2007, p. 309). News content in this temporal mode ranges from such mundane predictions as a weather forecast or the announcement of upcoming government action. But much of it is unique in its speculative orientation, leaving room for poetic license and creativity. It is in this speculative register, Neiger (2007, p. 311) notes, that:

> we meet journalists in their full might as creators of texts of 'reality' that has not yet occurred and may never come to be. That is, this 'reality' – which journalists report – happens only on the pages of the newspaper or on the air: from meetings that never take place, through strikes called off at the last moment, to chemical missiles that never leave their launching pad.

Journalistic writing about the future, then, presents an opportunity to discursively construct ways of dealing "with its darkest fears and wildest fantasies" (Neiger, 2007, p. 312). It is a kind of "prospective memory" that charts out future courses of action in the public good (Tenenboim-Weinblatt, 2013).

In the case technologies that constitute a possible existential threat, news coverage is powered by the twin impulses of protecting journalism as an institution and protecting society from evils, however speculative. As Fine (2007, p. 100) argues:

> Predicting and preparing for that which has not arrived is of considerable significance in permitting social order, giving confidence in coping with change … The vigorous dissemination of claims of what will transpire represents an attempt to create collective expectations – a shared future. We strive to narrow the range of possibilities … [thereby] creating the possibility of collective action.

The process of creating such collective expectations is achieved through narratives of news reporting as a "fact-centred discursive practice" (Chalaby, 1999). However, in the case of deepfake journalism, as we will argue, the storytelling is frequently based on a high level of speculation, which removes it from its anchoring in the verifiable and fact-based – even as this discourse is ironically concerned with verifiable fact.

Neiger (2007) distinguishes between four different levels of speculation about the future characterizing journalistic writing. Of these, the most speculative – and least used – discursive orientation is what he describes as "conjectured future" – a type of discourse based on speculating about an uncertain future, often one in which worst-case scenarios could become reality. Neiger notes, based on a study of headlines in the Israeli press over an 18-year period, that this discourse was infrequently used precisely because of its greater proximity "to science fiction than to purportedly factual reporting" (Neiger, 2007, p. 316). The stories we examine here are, in this sense, as much science fiction as they are the first

draft of history (Hyde, 2006). Labelling this conjectured future as science fiction is not meant to dispel risk. Instead, it highlights how this discourse combines possibilities enabled by technological advancement with an imagined future that has not yet occurred.

Neiger's examples of the news temporality of "conjectured future" often concerns society-wide political events (e.g. an upcoming breakout of armed conflict) that do not directly concern journalism. By contrast, the news discourse around deepfakes cannot be divorced from its implications for journalism. Whether explicitly stated or implied, speculation about deepfakes connects to a larger discourse of anxiety among journalists (Lowrey & Shan, 2018, p. 130). Writings about journalism, as Lowrey and Shan (2018) suggest, reveal "a path dependence strategy of projecting the known present into the unknown future" (p. 130). By analyzing journalistic discourses on deepfakes, we ask *how* these anxieties are expressed, *who* expresses them, and *what* they tell us about the profession's self-understanding and attempts at bolstering its authority.

Method

We examine journalistic discourses on deepfakes between January 1, 2018, when a GoogleTrends search demonstrates that the term first began to circulate, and July 1, 2019 when the US Congress held hearings on the technology. To carry out a comprehensive thematic analysis of coverage on the topic, we gathered a corpus of stories using the Nexis UK database. The database allows for systematic search of selected news sources and is widely used by Anglophone researchers, despite its well-documented gaps in coverage (e.g. Weaver & Bimber, 2008). A strength of this database is its international nature, which allows for a globally broad sample. We searched the database for all English-language news stories featuring three or more occurrences of "deep fake" or "deepfake," published in the time period between January 1, 2018 and July 1, 2019.[2] Our analysis excluded documents with

fewer than 500 words to focus on stories featuring in-depth discussion of the topic, encapsulating developed narratives about the phenomenon. This search resulted in a sample of 648 stories. After eliminating stories that were not relevant to the study,[3] duplicates, press releases, or those from student newspapers, the final sample consisted of 387 stories. These stories appeared in a global range of English-language outlets. While national legacy news outlets (such as the *Washington Post* in the US and the *Guardian* in the UK) predominated, the sample also included US regional newspapers (e.g. *Austin American-Statesman*), news websites and magazines (e.g. *Forbes*, *Time* and *HuffPost*), specialist tech publications (e.g. *Technology Review*) and transcripts of news shows from major US broadcasters (e.g. CNN). US outlets accounted for the majority of the sample, making up 226 of the total stories (58%), while UK outlets were the second-most frequent, represented with 71 stories (18%). Other countries with outlets represented in the sample included Australia, Canada, India, Malaysia New Zealand, Nigeria, Pakistan, Singapore, South Africa and United Arab Emirates.

The sample was subjected to multiple rounds of close reading by both authors with a focus on uncovering discourses around the social, political and journalistic consequences of deepfakes, allowing for the identification of key themes (Braun & Clarke, 2012; Fereday & Muir-Cochrane, 2006, pp. 82-83). In doing so we used combination of induction and deduction, as it common in thematic analysis (Braun & Clarke, 2012): While we drew on *induction* through a pre-determined conceptual focus on social, political and journalistic consequences of deepfakes, we took a *deductive* approach in allowing for the main themes to emerge through our close reading. Emerging themes were verified by both authors in the initial stages of reading, and the entire sample was subsequently coded according to these key themes, which are discussed below.

Though our sample represents a diverse range of stories about deepfakes, it is, like much journalism, largely event-driven and centered on a small number of key developments during the 18-month period covered by the study. While the initial coverage in January and early February 2018 focused on a Reddit group circulating an app for creating fake porn videos, the coverage swiftly turned to political consequences with Jordan Peele's Obama video and subsequent videos involving Mark Zuckerberg, Donald Trump, and Nancy Pelosi. Attention to these videos gave rise to detailed discussions of potential weaponization of deepfake technologies. Coverage intensified in mid-June 2019 as the Intelligence Committee of the US House of Representatives held a hearing on deepfakes and artificial intelligence, and finally circled round to the potential for the technology to be used for the manufacture of porn following on from the release of a "deepnude" app, turning images of clothed women into nudes, in late June 2019.

Findings

Our analysis suggests that journalistic coverage of deepfakes is highly speculative, and largely paints a dystopian picture of the consequences of the technology. Throughout our sample, only a small number of stories celebrate the creative potential of deepfakes as an innovative technology fostering artistic expression (for example, allowing filmmakers to insert deceased actors into movies). The vast majority adopt a much more pessimistic view. One area of concern arising in many stories throughout our sample period is the use of deepfakes to fabricate pornographic videos (see Maddocks, 2020). While this remains a serious potential abuse of the technology, our research focused on discourses about the civic and journalistic implications of deepfakes. Coverage of deepfakes envisions a dark future in which powerful "bad actors" weaponize the technology to undermine democratic societies. This future is premised on the ignorance and naivety of citizens who are unable to assess the

veracity of information. At the same time, however, this future may be salvaged by journalism and its associated practices, including fact-checking and holding the powerful to account. Below, we set out each of these key discursive categories in turn, analyzing what they tell us about journalism's place within the horizons of possibility for collective action:

(1) Fears of what *could* happen; worst-case scenarios: "bad actors" weaponize the technology to undermine democratic societies, and images and videos are no longer trusted;

(2) Assumptions embedded in worst-case scenarios of what *could* happen: citizens are ill equipped to exercise critical reasoning and media literacy, and therefore likely to fall prey to the deceptions of deepfakes;

(3) Debates over how to avoid what *could* happen: technological fixes, digital literacy, and the role of journalism as a vital institution.

Below, we discuss the relative prominence of each of these discourses. However, these are not mutually exclusive, but rather overlapping and interdependent. While the first – fears of what could happen – appear in some form in the vast majority of stories in our sample, the assumptions and debates arising from these fears articulate in diverse ways, dependent on the news events the stories are based on, and the contexts in which they appear.

(1) Fears of what *could* happen

Accounts of potential worst-case scenarios are central to the conjectured futures imagined in the coverage of deepfakes and appear in the majority of stories in our sample. Stories often call on readers to imagine a dark future resulting from the projected trajectory of the new technology. The conjectures are premised on fears associated with current social and political developments, including the intervention of foreign governments into US elections,

and the ease of spreading fake news through social media. The worst-case scenarios implied by coverage frequently involve the malicious interference of powerful "bad actors," with the Russian and Chinese governments, as well as Donald Trump himself, singled out as particularly plausible culprits:

> Deepfakes could [...] imperil the democratic process itself. It's not difficult to imagine a scandalous, but fake video being posted online right before polls open, or to imagine conspiracy theorists, or Trump himself, sharing a doctored video aimed at destroying a political opponent. (*The Statesman*, May 17, 2019)

These concerns, while salient throughout the sample, became particularly prominent in the light of the US House of Representative's Permanent Select Committee on Intelligence hearing on the challenge of AI and deepfakes, held on June 13, 2019. The committee, headed by high-profile Democratic representative Adam Schiff, called attention to the "nightmarish scenario" of resource-rich actors weaponizing the technology. Expert testimony complemented Schiff's alarmist narrative, suggesting that rogue nation states, including Russia, Iran and North Korea, may manufacture videos to disrupt US elections.

Articles on deepfakes openly advocated for fear of the technology in line with such dystopian visions. In Nigeria, the *Sun* wrote "Terrified yet? You should be" (February 25, 2018). Headlines included, "Fake videos: Everybody's a potential target" (*The Press*, January 5, 2019), "All of us stand to suffer" (*Scotland on Sunday*, February 13, 2019), and the horror-movie terror of the *Daily Star*: "Deep Fake video war: Point of no return for world just TWO YEARS away" (October 6, 2018). The *New York Times* portrayed deepfakes as marking a shifting epoch: "we find ourselves on the cusp of a new world -- one in which it will be impossible, literally, to tell what is real from what is invented" (October 17, 2018). *PC Magazine* suggested that deepfakes "could also open a Pandora's box of fraud, forgery, and propaganda" (June 7, 2018). In Australia, *The Age* considered the matter as already upon us,

“It might be too late to stuff the deepfake genie back in the bottle” (May 6, 2018). Not mincing words, a *Guardian* writer argued “we are sleepwalking towards a media dystopia in which nothing at all can be trusted” and that deepfake developers “are actively working, whether they realise it or not, to destroy liberal democracy” (February 18, 2019).

To explain the technology and provide informed speculation about deepfake technology, much of the reporting relied on computer science academics as authoritative news sources. A particularly prominent academic source is University of California Berkeley professor Hany Farid. Farid has long worked on technical solutions for identifying visual fabrications, first in still images and later in video. As an expert, he is regularly relied on to envisage negative consequences of the technology. For example, a *Quartz* opinion piece titled “The dystopian digital future of fake media” connected deepfakes with the end of democracy. His work was also featured in a *New Yorker* profile (November 12, 2018). Farid serves both as a source for explaining technical details and for speculating about their importance. For example, on CNN, he said “I don't know how you have a democracy if we can't agree on some basic facts of what's happening in the world around us” (January 31, 2019). Journalists draw on such experts to generate authoritative accounts, but these sources also have a vested interest, informed by the political economy of academic life, in promoting dystopian narratives to enhance their professional profiles.

Conjectured specificity and dystopian narratives

Many stories offering speculative reporting on future deepfakes problems conjured up remarkably specific visions of what could happen. Drawing on Neiger’s concept of “conjectured future,” we refer to this discursive practice as “conjectured specificity.” It is characterized by supplementing for the lack of actual examples of deepfake forgeries through imagining terrible scenarios that have not occurred. For example, in the House Intelligence

Committee hearing mentioned above, committee chair Adam Schiff offered a range of hypothetical uses:

> A state-backed actor creates a deep-fake video of a political candidate accepting a bribe with the goal of influencing an election. Or an individual hacker claims to have stolen audio of a private conversation between two world leaders, when in fact no such conversation took place. (The *Hindu*, June 23, 2019)

Other observers followed suit in concocting nightmare visions. The *Washington Post* suggested a deepfake "video could show a political leader advocating for the reverse of what she stands for, or portray bloody events that never happened. It could trigger riots, swing elections, and sow panic and despair" (September 7, 2018). In the UK, *The Times* asked, "Imagine what could happen if a deep fake audio recording of Donald Trump authorising a nuclear strike against North Korea was posted on social media websites?" (February 27, 2019). The *Irish Examiner* wrote: "Deepfake technology could take a bogus conspiracy theory – like 'Pizzagate', which falsely claimed that a Washington DC pizzeria was the centre of a child sex ring operated by Hillary Clinton and her campaign chairman – and make it appear even more credible, confusing the media and voters alike and sowing further discord among political parties, constituents, and even families" (November 23, 2018). Conjectured specificity makes technological developments understandable by grounding their potentialities in concrete-yet-fabricated examples. In doing so, they construct the imaginary for the technology as they explain it. This mode of news discourse differs from journalists' professed adherence to reporting on existing or past occurrences. Adopting this future-oriented temporality pushes past the representational grounds of journalistic authority by foregrounding judgment that something *could* happen.

In the near absence of *actual* examples of deepfake malfeasance in the political sphere, several news organizations concocted their own deepfakes to illustrate the terrifying

potential of the technology. Jordan Peele's video of Obama was a partnership with BuzzFeed News. In Australia, the ABC produced a deepfake video of Prime Minister Malcolm Turnbull in a lengthy report on the new technology (September 28, 2018). In the US, the *New York Times's* Kevin Roose conducted a face-swapping exercise (March 4, 2018), as did ABC News's *Nightline* (December 10, 2018). Meanwhile, in the UK, the *Mail on Sunday* created a fake video of actress Emma Thompson promoting Brexit (in contrast with her actual public position) as a means "to illustrate the more insidious powers of cutting-edge artificial intelligence (AI) technology" (June 16, 2019). These prominent news outlets took this additional step to showcase the potential harms of the technology in ways that didn't exist yet. While this may be a gimmick to gin up interest, it also raises questions about the consequences of emphasizing conjecture and potentialities over reality.

The disappearance of truth and the importance of visuality

Many examples point to specific instances of potential misuses of deepfake technology, but a commonly suggested consequence is the loss of belief in visual truth. While visual truth has always been problematic for journalism (Newton, 2001), deepfakes are associated with the passing of an era of visual belief (Vaccari & Chadwick, 2020). In the words of the *Washington Post*, "Now society will have to learn that video no longer guarantees reliability. Instead, it could be the biggest lie of all" (September 7, 2018). On ABC's *Nightline*, deepfakes were projected "to put our faith in visual media in jeopardy" (December 10, 2018). *The Guardian* went further to call it "a total corrosion of trust, not only in news media but in documentary evidence of many kinds" (February 18, 2019). The *Indian Express* noted, "it appears that the last medium that the viewer could regard without instinctive incredulity is compromised" (March 3, 2018). These authors both promote

existing visual journalism as trustworthy while imagining that this trust will quickly erode with deepfake technology.

Many discussions of visual truth placed deepfakes within a historical context. The *New Yorker*'s Joshua Rothman took the long view: "Historically, it will turn out that there was this weird time when people just assumed that photography and videography were true. And now that very short little period is fading" (November 12, 2018). Similar, the ABC in Australia commented, "we're used to turning on the TV or looking on Facebook and seeing footage of someone saying something and not having to question it. It's this trust that has allowed the rise of mass media broadcasters – even if people on TV said things you didn't agree with, you could at least be sure they were the people saying it" (September 27, 2018). Reports singled out the danger of undermining long-held beliefs about the authenticity of events that we can *see* and *hear.* In a widely cited post for the *Lawfare* blog, legal scholars Robert Chesney, Danielle Citron and Quinta Jerecic developed this argument in the context of social media:

> To be sure, defamation in American politics is nothing new. Hamilton and Jefferson played that game too. What is different today is that the falsehoods involve firsthand visual and audio "evidence" that our eyes and ears are deeply inclined to trust (not just written words that might more readily be dismissed), and the frauds can rapidly reach countless individuals. (May 29, 2019)

These analyses draw on an essentially biological explanation that stresses the primacy of multi-sensory experience. They embody concerns about the interplay of this unprecedented assault on our judgments of truth and the rise of social media that facilitate widespread sharing of unverified information.

Reports frequently linked the decline of trust in audio-visuals to a broader epistemological crisis of democracy, resulting in a breakdown of established authority. As

surveillance scholar Mark Andrejevic observed in *The Conversation*, deepfakes endanger "our ability to communicate truths to one another and to generate a consensus around them," undermining the public accountability vital to democratic societies (June 23, 2019). A *Daily Mail* article quoted Anthony Glees, Director of the Centre for Security and Intelligence Studies at the University of Buckingham:

> It's not just fake news, it's giving out fake news in the images of people who we believe. … This kind of Frankenstein technology will have a devastating impact on our politics. … The end result will be a collapse in the trust in political figures, which we need in order for our democracy to function. (June 16, 2019)

Conjectured worst-case scenarios, then, allowed journalists to generate a moral panic about something which *could* happen (but hasn't yet happened) – an imagined dystopian future in which the weaponization of deepfakes undermines democracy. This worst-case-scenario thinking underscores present anxieties around media technologies at a time of significant concern about the health of democratic institutions across the globe. It speaks to the speed with which social media have become objects of concern with the loss of control by journalists over channels of public communication, which requires turning our attention to assumptions in this discourse.

(2) Assumptions embedded in worst-case scenarios of what *could* happen

Journalistic narratives relied on a series of normative and common-sense assumptions in spelling out worst-case scenarios. Examining these assumptions tells us much about how the journalistic community understands its audiences and its relationship to social media. In line with the dystopian imagination characterizing the broader representation of deepfakes, these assumptions are underpinned by and give voice to journalism's anxieties about its authority as a social institution responsible for conveying factual information.

Journalists writing about deepfakes were frequently careful to specify that deepfakes did not emerge in a vacuum. A number of stories pointed to the continuity between longer-standing practices of misinformation, recent concerns about "fake news" and the emergence of deepfakes. For example, reporting suggested that "shallow fakes" (or the selective editing and clipping of real footage) have long served as a source of misinformation (e.g. *Observer*, June 16, 2019), but cautioned that deepfakes may be unique in their destructive potential (*The Hill*, June 5, 2019). Despite this cautionary approach, journalistic discourses are premised on assumptions of citizens as ill equipped to exercise critical reasoning and media literacy, and therefore likely to fall prey to deepfakes.

While the general narrative about deepfakes takes on a technological determinist position in suggesting that essentially agent-less technology drives human behavior, some commentators connected the problems of deepfakes to contextual elements. These included the growth of social media, political polarization, and a general penchant for motivated reasoning (i.e., privileging information that aligns with a preferred outcome) as a driver of gullibility. In a small number of cases, the discourse of conjectured specificity projected worst-case scenarios based on assumptions about citizens' limited ability to responsibly assess misinformation. A *Globe and Mail* columnist wrote, "Your dumbbell, Facebook-loving cousin just got a new weapon" (June 16, 2018). Similarly, *New York Times* technology columnist Kevin Roose argued, "People will share [deepfakes] when they're ideologically convenient and dismiss them when they're not. The dupes who fall for satirical stories from *The Onion* will be fooled by deepfakes, and the scrupulous people who care about the truth will find ways to detect and debunk them" (March 4, 2018). In South Africa, a writer for *Business Day* dismissed deepfake detection technologies, noting, "The problem is that it's all well and good that researchers and computers can detect the fakes, but ordinary people who could see those videos online may still be fooled" (September 19, 2018). These statements

establish a dichotomy between the reasoned examiner of information and the unthinking partisan-motivated social media user. The latter practices selective exposure, which the *Jerusalem Post* identified with partisan identities: "Many times, their main intent is to seek information that strengthens basic views they already hold" (January 25, 2019). *Al-Arab* Online noted, "Plagued by misinformation, rumour, canard and slander, social media are well able to feed biases and fuel communal, cultural and ideological tensions" (July 29, 2018).

The global resonance of these arguments suggests widespread fear that social media enable a hardening of political positions such that only like-minded information will get through, in line with widely circulating but increasingly discredited ideas of social media "filter bubbles" and "echo chambers" (e.g. Bruns, 2019; Pariser, 2011). In these estimations, the problem is not just the technology, but rather its interplay with processes of human cognition. In *The Atlantic,* Franklin Foer summed up these conditions: "human beings have displayed a near-infinite susceptibility to getting duped and conned-falling easily into worlds congenial to their own beliefs or self-image, regardless of how eccentric or flat-out wrong those beliefs may be" (April 8, 2018). Such explanations place deepfakes within the larger context of a swiftly changing media landscape and a decline of trust in institutions, while supporting a dark view of the capacity of citizens to navigate such complex changes. The construction of citizens as fundamentally naïve and ignorant is embedded in journalistic practices (e.g. Ekström & Tolson, 2017; Lewis et al., 2005) and has gained particular traction in recent years, in part driven by the rise of populist politics. It fuels the perception that citizens do not possess the digital literacy to assess the veracity of information – particularly as it circulates through social media – and that more substantive interventions from actors and institutions including governments, social media organizations, and journalism are therefore necessary to save democracy.

(3) Debates over how to avoid what *could* happen

The discourse of conjectured specificity is also prescriptive, with reporting addressing what should be done to avoid terrible things that *could* happen. Debates over possible solutions focused on technological fixes, the building of digital literacy, a need for proactive policy efforts, and the role of journalism in salvaging the future of democracy.

Technological fixes

One frequently mentioned weapon against deepfakes lies in detection technologies. A *Philadelphia Inquirer* writer noted that "The best hope for fighting computer-generated fake-porn videos might come from a surprising source: The artificial intelligence software itself" (January 6, 2019). A technological fix is seen as the answer to a technological problem, rendering questions about the political context irrelevant. At the same time, observers frequently noted that technological progress might outrun detection capacity. As one expert witness to the House Intelligence Committee hearing put it, detecting deepfakes is a race, or "like trying to monitor every bumblebee that's flying around America" (*Deseret Morning News*, June 14, 2019).

Social media's responsibility for detecting deepfakes

While deepfake technology was widely blamed, social media were frequently drawn in as the conduit through which deepfakes would be shared and have an impact. The discussion of deepfakes thus assumes the power of social media to dictate opinion, which is partly a reaction to the Brexit vote in the UK, the 2016 election of Donald Trump in the US, and populist political gains around the world. On MSNBC, the *New York Times's* Roose argued that social media "platforms have an enormous responsibility to make sure that they are not being used as funnels for this kind of misinformation" (May 10, 2018). Similarly, David Doermann, director of the Artificial Intelligence Institute at the University of Buffalo,

suggested that detection “also needs to be on the front end, not just after the images appear […] Pressure must be put on social media companies to realize that the way their platforms are being misused is unacceptable [and that] there should be warning labels on content that's not real or authentic whether that's determined by humans, machines or both” (*Deseret Morning News*, June 14, 2019). This approach speaks to larger concerns about social media in the realm of public communication, and connects to ongoing debates about the responsibility of platforms. Social media companies evade questions of accountability by differentiating themselves as forums hosting the content of others and therefore not liable to the expectations placed on publishers. Yet the emergence of fabricated content meant to do reputational or political harm pressures social media companies to better police content and develop socially beneficial solutions.

Enhancing digital literacy

Many commentators suggested that social media companies alone could not shoulder the responsibility for detection, and that media audiences will have to be more skeptical of content they encounter. A CNN business reporter summed up this position: “I think it's just a sort of matter of being skeptical of what you see online” (February 2, 2019). Also on CNN, US Senator Angus King echoed this suggestion: “The best defense is for people to be skeptical, to dig in and find out the facts.” This idea of skepticism even appeared in an editorial of a small US newspaper, the *Journal Times* of Racine, Wisconsin: “Don't believe everything you see, read or hear. Verify it, even if that ‘news’ is something that fits your own political or social views. Find another report from a different source before sending it along to all your like-minded friends and neighbors. If something seems preposterous, it often is. Don't be complicit in spreading misinformation by blasting it up and down your preferred information silo” (*Associated Press*, July 10, 2018).

Given these calls for skepticism, a *Wired* magazine reporter questioned the cognitive burden represented by the need for constant monitoring:

> Lost in the funhouse, we're told to be afraid – and to process every symbol we encounter with heightened diligence. The new catchphrase for web users is 'Verify, then trust.' That is, before you so much as laugh at a goony photo of someone, dig deep on URLs and metadata analysis, and scan for the ever-changing hallmarks of image manipulation and deepfakery. It's not enough to be defensive drivers on the information superhighway. We have to be prosecutorial ones. I have to admit, this all sounds very ... hard. (May 23, 2018)

Journalism's role in salvaging the future of democracy

Implicit (and sometimes explicit) in journalistic assessments of the difficulty of detecting deepfakes is the view that journalism continues to be vital as a professional institution dedicated to verification, in contrast to gatekeeperless social media. Writing about Israeli newspapers' reporting of conjectured futures, Neiger (2007, p. 317) found that journalism that engages in a very high level of speculation has significant political implications through its potential to create "social solidarity among the members of the community, who are driven by terror to seek refuge under the wings of politicians and generals." Much of the coverage we examined focused on the technology and its potential abuses – rather than journalism's response to it. However, when remedies for deepfakes were suggested, they frequently relied heavily on journalism and associated practices, including fact-checking and gatekeeping. *Global News Canada*, for example, cited misinformation expert Claire Wardle: "I think what we're recognizing now is societies need gatekeepers. … They need people who can be trusted, who can help us navigate the information ecosystem, and right now, we don't have that" (June 13, 2019). Other reports similarly emphasized the

responsibility of established media organizations, along with other traditional sources of institutional authority. According to the *Washington Post* (September 7, 2018):

> Quality media outlets need to emphasize how carefully they vet video. They should make sure their ethics codes and verification procedures adequately address the dangers. Otherwise, audiences will doubt any video including legitimate and important footage that media outlets gather in their own breaking news coverage and investigative work.

Similarly, an article in *The Atlantic* (April 8, 2018) argued: "Our best hope may be outsourcing the problem, restoring cultural authority to trusted validators with training and knowledge: newspapers, universities." Ultimately nightmare scenarios constructed through conjectured specificity rather than reporting on what *actually* has happened can only be resolved by resorting to journalistic authority, premised on practices of gatekeeping, fact-checking and responsible sourcing. Journalistic discourses on deepfakes support a future in which journalism is, once again, a vital and central social institution.

Conclusion

We have argued that stories about deepfakes are as much about what *might* happen in the future as they are about what has already occurred and what is happening in the present (Tenenboim-Weinblatt & Neiger, 2015). Given the lack of actually existing deepfake examples and resulting consequences, journalists operated in the mode of conjectured specificity to hypothesize and invent scary scenarios about the future. In the absence of examples, journalists went so far as to make up their own. This moves journalists away from the normative and epistemological foundations of their work as reporting on established facts to place them in the temporal position of inventing a future.

This temporal mode is revealing in several key ways. Journalists tend to highlight worst-case scenarios, heightening the danger of an unproven technology and instilling fears. Moreover, this fear gets connected to assumptions about publics who are too gullible to resist the falsity of deepfakes and other fabrications. Finally, the discourse on deepfakes gloms onto broader concerns about the unruliness of social media as public communication channels.

Journalists position deepfakes as one more reason not to trust social media content. The story of deepfakes told through conjectured specificity is of the growing encroachment of "bad actors" facilitated by a changing media ecology, particularly social media. These developments have both aided and worked in tandem with the rise of a new authoritarian politics as well as shifts in the tectonics of global power. Social media platforms aware of their reputational vulnerabilities have their own calculations to make in moderating and restricting content, as well as larger questions about their place in political communication. In all these ways, the journalistic discourse around deepfakes indicates social anxiety about the unrestricted flow of media content made possible by digital technology.

In constructing deepfakes as an arch-villain abetted by unruly social media, journalists promote the authority of professional journalism as a solution. Deepfakes represent an ideal topic for journalism's conjectured specificity precisely because the technology is easily portrayed as a scary story about profound threats to journalism and democracy – even if these threats are yet to be realized. Journalists, in turn, are well positioned to advocate for the value that authoritative professional practice to defeat this digital adversary. At a moment when the power of institutional journalism seems to be waning, the specter of deepfakes provides a rallying point for journalists to reassert their social value in both implicit and explicit ways.

This observation points to the need for journalism researchers to attend to uses of conjectured specificity. Deepfakes technology may be a particularly obvious and prominent

arch-villain. But it is certainly not the first nor the last technological development to represent a threat to the authority of journalism. If anything, new arch-villains are likely to arrive thick and fast in the polarized post-pandemic news environment. Investigating the imagined futures that journalists construct through the temporal mode of conjectured specificity can tell us much about how they see society and the role of journalistic institutions within it. Ultimately, it also points to the need to study whether and how the discursive mode of conjectured specificity operates in the context of other emerging technologies that might threaten to undermine journalism or the information environment more broadly. As a distinctive form of storytelling which breaks with established journalistic conventions, it requires a distinctive set of normative and epistemological standards. These can only be articulated on the basis of a detailed understanding of how journalists do their work of conjuring up the future across technologies, story types and social and political contexts.

References

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper (Ed.), *APA handbook of research methods in psychology* (Vol 2.) (pp. 57-71). Washington, DC: American Psychological Association.

Bruns, A. (2019). *Are filter bubbles real?* Cambridge, Polity Press.

Carlson, M. (2009). The reality of a fake image: News norms, photojournalistic craft, and Brian Walski's fabricated photograph. *Journalism Practice*, *3*(2), 125-139.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3), 416-431.

Carlson, M. (2018a). Fake news as an informational moral panic: the symbolic deviancy of social media during the 2016 US presidential election. *Information, Communication & Society*, ahead-of-print.

Carlson, M. (2018b). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society*, *20*(5), 1755-1772.

Chalaby, J. (1998). *The invention of journalism*. Houndsmills, Basingstoke: Palgrave Macmillan.

Clayton, J. (2021). Twitter boss: Trump ban is 'right' but 'dangerous.' *BBC News*, January 14.

Corner, J. (2017). Fake news, post-truth and media–political change. *Media, Culture & Society* 39(7), 1100–1107.

Creech, B., & Roessner, A. (2019). Declaring the value of truth: Progressive-era lessons for combatting fake news. *Journalism Practice*, *13*(3), 263-279.

Diakopoulos, N., & Johnson, D. (2019). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, online first.

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2020). Do

(microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics*, online first, doi: 10.1177/1940161220944364

Ekström, M., & Tolson, A. (2017). Citizens talking politics in the news. In *The Mediated Politics of Europe* (pp. 201-227). Basingstoke: Palgrave Macmillan.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, *5*(1), 80-92.

Finneman, T., & Thomas, R. J. (2018). A family of falsehoods: Deception, media hoaxes and fake news. *Newspaper Research Journal*, *39*(3), 350-361.

Grice, A. (2017). Fake news handed Brexiteers the referendum – and now they have no idea what they're doing. *The Independent*, January 18.

Gottfried, J. (2019). About three-quarters of Americans favor steps to restrict altered videos and images. Pew Research Center report, June 14.

Hao, K., & Heaven, W. D. (2020). The year deepfakes went mainstream. MIT Technology Review, December 24.

Hyde, J. (2006). News coverage of genetic cloning: When science journalism becomes future-oriented speculation. *Journal of Communication Inquiry*, *30*(3), 229-250.

Jack, C. (2017). Lexicon of lies: Terms for problematic information. Data & Society Research Institute report.

Johnson, B. G., & Kelling, K. (2018). Placing Facebook: "Trending,""Napalm Girl,""fake news" and journalistic boundary work. *Journalism Practice*, *12*(7), 817-833.

Kavanagh, J., & Rich, M. D. (2018). Truth decay: A threat to policymaking and democracy. Rand Corporation Briefing Paper.

Lewis, J., Inthorn, S., & Wahl-Jorgensen, K. (2005). *Citizens or consumers? Buckingham: Open University Press.*

Lowrey, W., & Shan, Z. (2018). Journalism's fortune tellers: Constructing the future of news. *Journalism*, *19*(2), 129-145.

Maddocks, S. (2020). 'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political'deep fakes. *Porn Studies*, online first.

McCarthy, C. (2019). Deepfakes could be good news. *Spectator USA*, February 2.

McNair, B. (1998). *The sociology of journalism*. London: Arnold.

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, *54*(1), 1-41.

Neiger, M. (2007). Media oracles: The cultural significance and political import of news referring to future events. *Journalism*, *8*(3), 309-321.

Newton, J. (2013). *The burden of visual truth: The role of photojournalism in mediating reality.* Mahwah, NJ: Lawrence Erlbaum Associates.

Nielsen, R.K,, & Ganter, S. A. (2018). Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New media & society*, *20*(4), 1600-1617.

Pariser, E. (2011). *The filter bubble*. London: Penguin UK.

Phillips, T. (2018). Bolsonaro business backers accused of illegal Whatsapp fake news campaign. *The Guardian*, October 18.

Schwartz, A. B. (2015). *Broadcast hysteria.* New York: Hill and Wang.

Silverman, C. (2016). This analysis shows how viral fake election news stories outperformed real news on Facebook. *BuzzFeed*, November 16.

Tamul, D. J., & Martínez-Carrillo, N. I. (2018). Ample sample? An examination of the representativeness of themes between sampling durations generated from keyword searches for 12 months of immigration news From LexisNexis and newspaper websites. *Journalism & Mass Communication Quarterly*, *95*(1), 96-121.

Tandoc Jr, E. C., Jenkins, J., & Craft, S. (2019). Fake news as a critical incident in journalism. *Journalism Practice*, *13*(6), 673-689.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital Journalism*, *6*(2), 137-153.

Tenenboim-Weinblatt, K. (2013). Bridging collective memories and public agendas: Toward a theory of mediated prospective memory. *Communication Theory*, *23*(2), 91-111.

Tenenboim-Weinblatt, K., & Neiger, M. (2015). Print is future, online is past: Cross-media analysis of temporal orientations in the news. *Communication Research*, *42*(8), 1047-1067.

Thornton, B. (2000). The moon hoax: Debates about ethics in 1835 New York newspapers. *Journal of Mass Media Ethics*, *15*(2), 89-100.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society*, *6*(1), 1-13.

Vos, T. P., & Thomas, R. J. (2018). The discursive construction of journalistic authority in a post-truth age. *Journalism Studies*, *19*(13), 2001-2010.

Waisbord, S. (2018). Truth is what happens to news. *Journalism Studies* 19(13), 1866-1878

Ward, S. J. (2009). Journalism ethics. In K. Wahl-Jorgensen and T. Hanitzsch (Eds.), *Handbook of Journalism Studies* (pp. 295-309). New York and London: Routledge.

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe report, DGI (2017)*, *9*.

[1] Note that there is a growing literature in computer science on automated detection of deepfakes, suggesting that there are systematic ways of determining whether audio-visual material is "authentic" or "fabricated" (e.g., Mirsky & Lee, 2021)

[2] We arrived at this search strategy after several rounds of piloting search terms and occurrences, following the recommendations of Tamul and Martinez-Carillo (2011).

[3] Most stories in this category were highly technical accounts, often from technology-focused websites and magazines, of how deepfakes are manufactured. As such, they did not engage substantively with the social and political implications of deepfakes.