2020

Politics and porn: how news media characterizes problems presented by deepfakes, Critical Studies in Media Communication

Chandell E. Gosse
Western University, cgosse@uwo.ca

Jacquelyn Burkell
The University of Western Ontario

Politics and porn: how news media characterizes problems presented by deepfakes

Chandell Gosse & Jacquelyn Burkell

View supplementary material ⬚

Published online: 24 Oct 2020.

Submit your article to this journal ⬚

View related articles ⬚

View Crossmark data ⬚

Check for updates

Politics and porn: how news media characterizes problems presented by deepfakes

Chandell Gosse ⬤ and Jacquelyn Burkell

Faculty of Information and Media Studies, Western University, London, Canada

ABSTRACT
"Deepfake" is a form of machine learning that creates fake videos by superimposing the face of one person on to the body of another in a new video. The technology has been used to create non-consensual fake pornography and sexual imagery, but there is concern that it will soon be used for politically nefarious ends. This study seeks to understand how the news media has characterized the problem(s) presented by deepfakes. We used discourse analysis to examine news articles about deepfakes, finding that news media discuss the problems of deepfakes in four ways: as (too) easily produced and distributed; as creating false beliefs; as undermining the political process; and as non-consensual sexual content. We provide an overview of how news media position each problem followed by a discussion about the varying degrees of emphasis given to each problem and the implications this has for the public's perception and construction of deepfakes.

Introduction

Deepfakes—fake videos that use artificial intelligence (AI) to place one person's face (learned from a set of existing images and video, known as a faceset) on to another person's body (known as a donor body)—were first adopted to create non-consensual sexual videos. Early productions were posted to Reddit and several pornography sites online. These spaces celebrated sexual deepfakes and contributed to a "broader environment in which women's images are understood as consumable, malleable, and brought into being for the enjoyment and gratification of men" (van der Nagel, 2020, p. 3). Early press coverage by *Motherboard* writer Samantha Cole (2017, 2018a, 2018b, 2018c, 2018d) similarly focused on non-consensual sexual deepfakes and the associated issues of the objectification and violation of women in such videos (Cole, 2017). Deepfakes can be used for a variety of purposes, including entertainment, non-consensual pornography, and political misinformation, but news media typically focus on political uses (Maddocks, 2020; van der Nagel, 2020). In this study, we use discourse analysis to understand how news media position the problems posed by deepfakes and to assess how these problems were framed in relation to one another in the earliest days of this technology.

CONTACT Chandell Gosse ✉ cgosse@uwo.ca 🖃 FIMS & Nursing Building, Room 4056 London, Ontario, Canada, N6A 5B9

We identify four problems: deepfakes are easily produced and easily distributed; deepfakes promote false beliefs; deepfakes undermine the political process; and deepfakes are used to make non-consensual sexual content. The goal of this paper is to examine these problems to understand how deepfakes are presented to the public.

Cultural context

Deepfakes is a new technology, first used for a familiar purpose: to objectify and demean women. Misogyny is not, however, "built in" to the technology; instead, the decision to use the technology to create sexual deepfakes rests with the users and reflects the misogynistic culture within which the technology is deployed (Burkell & Gosse, 2019; Vickery & Everbach, 2018). In fact, fake porn—or, non-consensual sexual imagery—has long circulated in the public arena, and particularly in online spaces (Burkell & Gosse, 2019; Maddocks, 2020). Sexual deepfakes, however, is the first form of fake porn to receive much attention in the news media. Sexual deepfakes are one of many ways in which women are objectified in digital and visual culture (van der Nagel, 2020). They also strip away the agency and autonomy of the women involved, including both the person whose face is placed into the video and the adult performer whose body "is erased, edited, and recirculated" (Maddocks, 2020, p. 2).

Mis/using images of women has been the subject of scholarly discussion since the 1970s, when scholars began focusing on the problematic depiction of women in media (Kilbourne, 1979) and the overwhelming tendency to treat women as subjects to-be-looked-at rather than subjects who do the looking (Mulvey, 1975). These critiques, and countless others, illuminate the limited and sexualized representation of women in North American popular culture. The use of deepfakes to make non-consensual sexual content is intrinsically tied to this culture (van der Nagel, 2020), but is also part of a larger trend that hijacks women's agency and privacy by using images—real or fabricated—to humiliate and disempower them (McGlynn et al., 2017; Powell et al., 2018). As van der Nagel (2020) argues, "By treating women's faces as a digital resource to be edited onto sexual bodies by artificial intelligence, [deepfakes] reinforces the idea that women exist as sexual objects" (p. 3).

The mis/use of women's bodies in digital media is embedded within digital and visual culture itself. Since its release in 1990, Photoshop and related image or video altering software has been used to create sexualized images of women (McGlynn et al., 2017). Over the last decade women's digital intimate images in particular have been weaponized and used to inflict harm. Collectively known as image-based (sexual) abuse, these practices include, among others, "relationship retribution," "sextortion," "sexual voyeurism," "sexploitation," and "sexual assault" (Powell et al., 2018, p. 306). Of these practices, the non-consensual distribution of intimate images, colloquially known as revenge porn, has received the most attention (Bates, 2017; Citron, 2014). For the most part, image abuse relies on having access to intimate images of the woman against whom the abuse is perpetrated. Deepfakes, however, undercuts that reliance, offering the possibility of perpetrating image-based abuse without access to actual intimate photographs or videos of the woman who is targeted. Using deepfakes, *any* woman can be made to appear in pornography: all that is required is a sufficient corpus of facial images, an appropriate pornographic video into which the woman's face is to be placed, some technical know-how, and deepfakes become one more, particularly powerful, way women's bodies can be used

against them. In this paper, we focus on women for two reasons. First, because that is the language and approach of the articles in our data, and second, because there is no research or data to understand how sexual deepfakes impact gender non-binary or transgender individuals. However, given what we know about online abuse more broadly, which is that BIPOC, gender non-binary, and members of LGBT communities are disproportionately targeted, we believe that there needs to be a specific focus on the heightened vulnerability of equality and equity-deserving groups.

Deepfakes, and related software such as FakeApp and DeepNude, democratizes a capacity that was historically restricted to experts in the entertainment industry: namely, the ability to create convincing fabricated video of people doing and saying things that did not happen in real life. Deepfakes remain somewhat difficult to produce, but as the technology grows more sophisticated the potential for more people to create them becomes increasingly likely (see Burkell & Gosse, 2019). Since the initial release and subsequent widespread use of the technology, deepfake productions have been recognized for their potential to mislead audiences, especially when combined with other applications such as realistic voice altering software. Although sexual deepfakes remains the most common application (Ajder et al., 2019; Chesney & Citron, 2019), deepfakes have already been used for other ends, such as creating false videos of celebrities (e.g. a series of Nicolas Cage videos; O'Reilly, 2018), and politicians (e.g. the fake Barack Obama video produced by Jordan Peele and Buzzfeed; BuzzFeedVideo, 2018). These examples differ from the intention behind sexual deepfakes as they were created as a form of entertainment and to demonstrate the power of deepfake technology. This is very different from the intention of sexual deepfakes—which admittedly varies, but objectively centers on demeaning, objectifying, and humiliating women.

Concerns about the production and circulation of false videos have naturally led to a focus on content authentication and deepfake detection/identification. Researchers have been working toward creating detection techniques in order to verify and authenticate content in order to minimize the potential negative impact of these productions; however, even as new detection/authentication approaches are identified, producers of the false videos are developing techniques to defeat them. For example, in 2018 researchers discovered one way to detect deepfakes: blinking (Li et al., 2018). Soon after, however, higher rates of blinking became incorporated into new models of deepfakes, rendering this detection technique useless. The result is, as Day (2019) puts it, a "computational arms race … between coders who generate fakes and those who detect them" (p. 108).

Other, non-technical, solutions to the problem of deepfakes are also being explored. Wagner and Blewer (2019) suggest that society also requires an understanding of the sociopolitical nature of the problem. Specifically, they argue for a combination of visual information literacy and a feminist approach to AI. They believe that "[u]nderstanding the material implications of the hyperreality of deepfakes might be how one gets over their becoming normalized" (p. 42). Part of this materiality is the unequal distribution of who is seen and who does the seeing; there is a strong gender disparity between who is impacted by deepfakes and who is not as women are consistently coded into sexual deepfakes for the pleasure of others. For Wagner and Blewer (2019), critically minded and feminist approaches to AI, in combination with preparing "persons to see things with a lens that is not merely critical, but oriented towards a notion of social justice" (p. 42), can potentially interrupt the trend of creating sexist and misogynistic deepfakes.

One of the key concerns about deepfakes is the presentation of something "unreal" as "real." This concern operates in the realm of political disinformation, and also in the realm of sexual deepfakes—but with sexual deepfakes there is the additional (and perhaps more critical) issue of the representation itself. The problem of disinformation can be addressed by more effective detection and verification techniques, coupled with increased education of the viewing public to equip them to identify and discount "fake news." Detection, verification, and even removal of sexual deepfakes, however, will not mitigate the harms experienced by the women who are targeted. Detection and removal are certainly important in this context, but there is a level of reputational damage and a cognitive burden associated with someone else seeing, or you seeing yourself, in a sexually compromising film or image that cannot be undone by detection and removal strategies. For the women who have their faces used in sexual deepfakes, the "damage has already been done" (Cox, 2018): detecting and removing the content will help to contain, but will certainly not eliminate, the harm caused by these productions.

In the case of both sexual and political deepfakes, the technology stands to alter, and indeed threaten, the cultural fabric. Public understanding of this technology is critical for effective responses, and for effective participation in policy-making (Durant, 1999). Academic discussion of deepfakes is limited—though that is changing—partly because they "are simply too recent for anyone [in academia] to have much perspective" (Fletcher, 2018, p. 457). As a result, the harms associated with deepfakes have been largely constructed by news media. As such, this coverage serves not only to educate the public, but also to shape public understanding, and as such, knowing how problems associated with deepfakes are framed becomes incredibly important.

Methods

Conceptual framework

Discourse analysis is a traditional and important method used in media and cultural studies to help understand the way language and text shape systems of social meaning (Fairclough, 1989; Tonkiss, 1998). Using a combination of critical and feminist discourse analysis, we report on the way power relations unfold through the use of language and the position of text. In particular, we critique the way power—especially patriarchal power—naturalizes the treatment of certain kinds of harms. This naturalization occurs through the subtle and sometimes flagrant expression of beliefs and values that, through repetition, appear as though they are "common sense" (Fairclough, 1989, p. 77). Once they appear as common sense, it is easier for discursive practices that replicate power imbalances to go undetected and unchallenged.

Discourse analysis also provides a great deal of freedom to think critically about the phenomenon under study. As a method, it does not pretend to be objective because the data themselves are not objective, and the motivation for applying the framework is overtly political. Feminist critical discourse analysis, in particular, is part of "an emancipatory critical social science which […] is openly committed to the achievement of a just social order through a critique of discourse" (Lazar, 2007, p. 145). Drawing on Lather (1986, p. 259), Lazar (2007) argues that feminist critical discourse analysis "is scholarship that makes its biases part of its argument" (p. 146).

In discourse analysis the reasons for being interested in a particular topic become critical parts of the analysis (Johannesson, 2010). In this study, a concern over how deepfakes are being presented to audiences of news media is the central concern. Using a feminist critical discourse lens, we investigate how problems with technology are framed for public consumption, examine who those problems impact, and highlight whose experiences are left out.

Data collection, sample & analysis of news coverage

We used Factiva, a news media search engine, to gather articles with the terms "deepfake/s" and/or "fakeapp" in the headline or body. We used the term deepfake/s for obvious reasons, and added the term fakeapp because it was the name of an application created around the time of data collection that helped automate the creation of deepfakes. Factiva provided us with a corpus of articles drawn from English language major publications from the United States, Canada, and the United Kingdom, such as the Washington Post and the New Yorker, the Globe and Mail and the Toronto Star, and the Guardian and BBC, as well as regional newspapers that circulate offline and/or online. Overall, our sample represents news media from a variety of larger, well-known publications, as well as smaller scale, local publications.

Results were limited to English language articles published between 1 December 2017 (shortly before news of deepfakes broke) and 31 October 2018 (when we began data analysis). The initial search returned 325 articles. Close duplicates (i.e. articles by the same author, using the same sources and quotes) were removed through a combination of automatic filtering provided by Factiva and manual review. Articles not topically related to deepfakes and compilations of weekly news that tangentially mentioned deepfakes were also removed. These filtering processes resulted in a final sample of 123 unique articles, all addressing deepfakes or fakeapp.

We then used Nvivo 12, a qualitative analysis software, to navigate data analysis. First, we coded every problem discussed in our dataset. Second, we identified the primary focus and main problem for each article. Third, we grouped main problems according to themes. At this stage both authors and a third research assistant grouped these codes together independently. We then discussed our themes, reconciled any discrepancies, and finalized our final four themes. Last, we went back to the original coded problems in the first step and organized them according to our four themes. This stage was also done independently by three coders, followed by another process of discussing, refining, and finalizing. Nvivo 12 also helped us keep track of discursive tendencies we noticed in the sample (as discussed below under *Hierarchy of Harm*). In the following section, we outline what *news media tell the public* about the problems associated with deepfakes. We intentionally did not speculate or expand on the problems. As such, the findings below are a reflection of the data, separate from how we view the problems.

Findings

Although deepfake technology can be used, as one story puts it, for "good and evil" (Article 91), in general the press coverage we sampled identified deepfakes as problematic with potential negative consequences. A small subset of article discussed the humorous

use of deepfakes, like placing Nicholas Cage's face in films like Lord of the Rings and Indiana Jones, and the potential of this technology for use by Hollywood. Overwhelmingly, however, news coverage positioned the problems and consequences associated with deepfakes in four partially overlapping ways. The first identified problem focused on how easy it is to create and share false video content using deepfakes and applications like FakeApp. The second was the possibility that deepfakes could lead audiences to false beliefs and disrupt any semblance of a shared framework for understanding reality. The third problem was that deepfakes could undermine the political process in the form of disinformation and fake news. And the fourth problem was the use of deepfakes to create non-consensual sexual videos.

Problem one: deepfakes are easily produced and easily distributed

The first identified problem in the press coverage reflected a widespread belief that deepfakes can be easily produced and distributed. This belief hinged on three conditions that, together, magnify the potential consequences of deepfakes: accessibility of software; availability of data; and an environment that lacks verification of shared content.

The first condition focused on the accessibility of the technology used to create deepfakes. FakeApp, an application that allows users to easily create deepfakes, became available online in early 2018 to anyone with a desktop computer. The program basically automated the creation of deepfakes. As one article noted, "the New York Times reported on technology available for download on the internet that enables ordinary people with access to computer networks to create fake videos that are nearly undetectable as fake" (Article, 118). The interesting part of this positioning is the use of the word *ordinary*, which indicates and underlines the fear associated with a powerful piece of new technology falling into the hands of average online users. The coverage also points out that the ease of production will only increase as the technology inevitably and quickly improves: " … FakeApp and its ilk are getting better and easier to use by the day. During our month-long testing of FakeApp it has evolved from a jury-rigged bundle of command-line utilities to a fairly slick one-button application" (Article 42).

The second condition rests on the fact that the production of deepfakes requires large quantities of images of the individual being *swapped into* or *superimposed on to* (note: neither of these terms are entirely accurate) the fabricated video, and that such quantities are readily available. Press coverage pointed out that there is an abundance of available content of celebrities and politicians that can be used to create realistic deepfakes. While public figures are obviously accessible targets, articles also reminded readers that social media creates optimal conditions for the potential creation of deepfakes that feature anyone. As one article notes, "[…] people today are constantly uploading photos of themselves to various social media platforms, meaning someone could use such a technique to harass someone they know" (Article 61). Another article states:

> Recently, the anonymous administrator from a deepfake porn site announced that they want to "democratise" the practice of deepfaking—making it easier than ever to splice people you know into explicit porn films … All anyone needs to create fake porn is a selfie or video of their target—*and there's more than enough of those floating around the web*. (Article 3, emphasis added)

Discussion of the third condition, that digital environments lack verification of shared content, was less obvious than the previous two conditions—though nonetheless present—and strongly intersects with the other problems identified in this paper. This condition emphasizes a lack of careful attention by users to the sources of information shared online as well as the ability to easily share content, which means that deepfakes can "quickly go viral online" (Article 88). It thus cautions that because unverified content can be easily shared, deepfakes can and will become a "very powerful way to disseminate credible disinformation to the world" (Article 58).

By writing about deepfakes using these conditions, news media constructs an idea of harm that points to problems outside of the technology itself. Things like the environments in which deepfakes exist, which is one that encourages cursory engagement through likes and shares, and the accessibility and availability of the necessary software and data, impresses on readers that the problems presented by deepfakes are a problem with the digital ecosystem. This moves responsibility away from people who choose to use the technology for nefarious purposes, and instead toward less tangible by-products of digital environments. There is even an element of "victim-blaming" in this discourse, which focuses on the availability of images (in many cases uploaded by "everyday" users into social media profiles) as part of the problem.

Problem two: deepfakes promote false beliefs

The second problem identified in the press coverage was that the technology "will eventually fool even the sharpest eyes" (Article 77), and that viewers might come to hold false beliefs as a result of fake content. Current deepfakes are far from perfect, and many can be easily distinguished from actual video content. There is no doubt, however, the technology will improve over time, with the eventual result that false videos will be indistinguishable from true recordings. As one writer notes: "Photo manipulation technology long ago took away still photos' value as incontrovertible proof of something. Now we can add fake videos that, like the computergenerated graphics in motion pictures, can make the unreal seem completely plausible" (Article 118). As one article states, sophisticated video altering software presents socio-political concerns, including the potential "for police bodycam footage to be tampered with" (Article 112).

The obvious problem that false videos could lead to specific false beliefs was related, in much of the coverage, to the larger concern that false videos could undermine our shared reality and agreement on basic facts—both by representing as "true" that which is not, and by leaving audiences unsure what, if any, content can be trusted. As one reporter explained, "[…] with the looming horizon of computer generated 'deepfakes' [..] the line between reality and fiction is blurring" (Article 32). Commentary reflecting this issue includes:

> No voice is safe. No image is safe. Incorporate AR (artificial reality) techniques into this "end of the truth" scenario, and we simply will not be able to tell reality from fiction. If you can't believe your eyes and your ears, what can you believe? (Article 76).

And similarly: "If you can't tell a fake from reality, then it becomes easy to question the authenticity of anything" (Article 55). One article offered a particularly pithy summary: "Seeing isn't believing anymore" (Article 65).

The overall concern here is that deepfakes accelerate an already eroding trust in the things we see, hear, and read. In so doing, they blur the lines of reality and cause damage to the audiences who view them. While the issue of false beliefs was prominent enough to be considered a problem in itself, a concern over political content (read: politicians *saying* things they did not actually say, propaganda, etc.) was typically associated with this problem (and thus is deeply connected to problem three, noted below).

Problem three: deepfakes undermine the political process

The third problem focused on the potential political fallout of false videos. The news articles raised the concern that this could undermine the electoral process, threaten national security, and operate as an extension of the ongoing issue of fake news and disinformation. The strong focus on this concern demonstrated in these articles might be partly explained by considering other news headlines circulating at the time of data collection. In the timeframe for which we collected data there was increased concern and coverage about alleged Russian interference in the 2016 US federal elections (Galante & Shaun, 2018), and the articles focused on the possible implications for future elections. For example, one article notes that "[researchers are] betting whether or not someone will create a so-called Deepfake video about a political candidate that receives more than 2 million views before getting debunked by the end of 2018" (Article 47; spoiler: there was no such video).

Concerns about the impact of synthetic video on the electoral process were rooted in the notion that these productions could create false beliefs in audiences. Deepfake productions were discussed as "videos that can appear to present a person, even a politician, saying or doing something they never actually did" (Article 92). The coverage includes many evocative examples: one story, for example, states "Any given person can now create propaganda on their political enemies. President Trump's face can be superimposed into a video of him doing cocaine, or a young Bernie Sanders can be shown in KKK rallies … " (Article 91).

The concern that synthesized video could undermine our shared sense of reality was itself viewed as a specific threat to democracy: "'Democracy assumes that its citizens share the same reality,' an op-ed concluded, 'We're about to find out whether democracy can be preserved when this assumption no longer holds'" (Article 70).

Readers were also cautioned that disinformation in the form of synthesized video could lead to "[a] misinformed public, moral outrage, and greater polarization" (Olson, 2018). Deepfakes were positioned as being used potentially by "Russian foes in disinformation campaigns" (Murphy, 2018) for "political sabotage and propaganda" (Article 118), and as having the potential to "stir up conflict between nations or communities" (Article 74) since they could, for example, "be used to produce convincing video of Donald Trump making abusive remarks about Muslims or Vladimir Putin declaring war on Britain" (Article 74). One article quoted Edward Lucas, senior vice president of the Centre for European Policy Analysis, as saying: "We're rapidly moving into an era where the Russians, or any other adversary, can create our public figures saying or doing things that are disgraceful or highly corrosive to public trust … And we're not remotely ready for this" (Article 42).

The need for better detection—identifying false or synthesized videos—was often flagged as a critical response to political or security concerns. This issue arose in the press coverage in relation to reports of funding by the United States Department of Defense for a project focused on the detection of fake videos. For example:

> The Defence Department's Darpa [*sic*] agency wants machines that can adapt to their environment and are better at spotting counterfeit imagery. The US military is backing artificial intelligence research into areas including systems with "common sense" and networks that can spot images manufactured by other AIs (Article 14).

The coverage focused primarily on the difficulty of detecting these videos; for example, one article noted that "the kind of tech Darpa [*sic*] is pinning its hopes on is 'decades away'" (Article 27).

The articles in our sample seemed to construct a fear of deepfakes as the "ultimate" threat to a democratic society. While fake news and disinformation *do* pose risks to audiences, and this is a legitimate concern to have, the coverage that took this angle relied primarily on hypothetical scenarios, possibly stoking fears over deepfake uses that have not yet been realized. Absent from the coverage is any discussion over how sexual deepfakes can also dupe voters and undermined political processes "just as much as any digitally manipulated political speech" (Maddocks, 2020, p. 4).

Problem four: deepfakes are used to make non-consensual sexual content

The fourth problem centered on the use of deepfakes to create non-consensual sexual deepfakes. In the articles where this issue was more than just tangentially mentioned, this issue was discussed in two ways: social media companies' responses to such content and the harms such content poses to women.

Much of the discussion around sexual deepfakes centered on reports that websites like PornHub, Reddit, and Tumblr, among others, were publicly decrying sexual deepfakes and announcing their efforts to remove and ban them from their websites. "Twitter, Reddit and Pornhub have all become the latest top web platforms to ban AI-generated pornography, also known as 'deepfakes,' after terming it as non-consensual porn" (Article 80). These articles highlighted the misogynistic intent for which deepfakes are being used, which certainly raises awareness that this is a central issue within the scope of deepfake technology, but the main point was to notify readers of social media companies' responses.

Broadly speaking, there were three harms to women mentioned in the press coverage: emotional and psychological distress; loss of autonomy to one's body and one's reputation; and the connection between deepfakes and other possible crimes.

While there was some discussion of the emotional and psychological harm caused by deepfakes, for example, explaining that this type of victimization can leave victims "feeling distressed and humiliated" (Article 3), this coverage typically extrapolated from more well-known cases of cybermisogyny, such as non-consensual distribution of intimate images. For example, one article quoted a woman who explains: "The aftermath [of having intimate images non-consensually shared] is worse, not knowing who will see it and who has already. Your life becomes a constant worry" (Article 3).

News media also discussed the lack of consent involved in deepfakes. While one article describes deepfakes as "violat[ing] bodily autonomy on an industrial scale" (Article 42),

most articles limited their discussion of consent to simply pointing out that pornographic deepfakes are created without the consent of the women in them. Specifically, the articles focused on the women whose faces were placed in the videos, not the adult performers whose bodies were also violated. Relatedly, there was some concern raised over the implications deepfakes have for women's reputations: "Some deepfake videos may create false statements of fact about a person's presence and actions that lead to a loss of reputation of that person" (Article 54). In most cases, however, the focus of concern was directed toward celebrities, in part because deepfakes targeting celebrities gained the most notoriety. For example, one article discussing Emma Watson and Gal Gadot explains: "There is an obvious problem here—it deals with concern about their image and the violation that they experience by the implication they are involved in such a film" (Article 91).

Lastly, there was a concern that targets of deepfakes might become victims of abuse and identifiable crimes such as harassment, defamation, blackmail and extortion. In this way deepfakes was often aligned with other types of online abuse, and image-based abuse in particular. For example, during the period covered in our data, the United Kingdom (UK) had recently criminalized "upskirting," or the act of surreptitiously taking photos underneath a woman's skirt. In the press coverage, the new ban on upskirting was connected to wider trends in cybermisogyny and used to highlight the difficulty of developing legal responses to problems associated with new technology, many of which disproportionately impact of equality and equity-deserving groups.

> "Online abuse is the focus of significant feminist activism these days, as social media platforms flail ineffectually against a torrent of cyber-misogyny." The law has struggled to keep up with technology, in particular around image-based abuses. "Deepfake" porn—producing fake pornographic images based on pictures of real people—is the latest tech-enabled sexual abuse to attract demands for a ban (Article 15).

Disappointingly, the objectification of women and misogynistic intent of sexual deepfakes was rarely, if at all, mentioned in our sample. One exception is an article that mentioned the degradation that would obviously come as a result of sexual deepfakes: "In many cases these videos are pornographic, some containing celebrities, and others featuring people known to the creator. They are all equally degrading" (Article 43).

A hierarchy of harms

In many of the articles, sexual deepfakes served the role of "origin story," providing a segue into discussions of other consequences. The following quote is typical of this treatment:

> Predictably, faceswapping tech was quickly hijacked by the darker corners of the internet, in particular for inserting celebrities' faces on to actors in pornographic films. There are also fears it could be used to spread political disinformation on social media, potentially influencing the outcome of elections. As people become wise to fake news stories, will they start to be fooled instead by fake videos? Imagine the inflammatory lies that could be added to the lips of politicians (Article 23).

While it is important to trace the history to provide context for the discussions of deepfakes, the fact that sexual deepfakes is framed as a *background* story is a problem. In 2019 the startup DeepTrace—whose primary objective is to research "deepfakes evolving

capabilities and threats" and provide "crucial intelligence for enhancing our detection technology"—conducted the most comprehensive study of deepfakes to date (Ajder et al., 2019). Analyzing 14,678 deepfake videos, which they admit is not an exhaustive sample, they reported that 96% of their dataset was pornographic (Ajder et al., 2019). In other words, the issue of sexual deepfakes is not a *background* issue—it remains overwhelmingly true that to talk about deepfakes is to talk about fake pornography.

Even when deepfakes pornography is not relegated to the position of "background" or "origin story," this use of the technology is often positioned as less important than the use of deepfakes to promulgate false information:

> Late last year, so-called "deepfake" pornographic videos began to surface online, with celebrity faces realistically melded to different bodies. "It happened in the regime of pornography rather than propaganda," said Jack Clark, head of policy at OpenAI … "But nothing about deepfakes suggests it can't be applied to propaganda" (Article 4).

In this quote, Clark seems to be subtly suggesting that an application of deepfakes to propaganda would be more serious than its use in sexual deepfakes. The same suggestion is made in another article:

> Deepfakes, as they are known, are videos that use machine learning to superimpose one person's face into the video of another person. It has been used to make fake pornographic videos of celebrities, and with the right editing, it can be entirely convincing. However, there is an *even more nefariou*s use for the technology. Foreign powers or even domestic troublemakers could use the techniques to create propaganda or synthesized events (Article 54, emphasis added).

Another article is more explicit in subordinating the harms experienced by *targets* of deepfakes (including porn) to the harms experienced by those who *view* fake news:

> Who are the victims of deepfakes? Is it the women who've been blackmailed with nonconsensual and completely fabricated revenge porn videos, their faces stitched onto pornstars' bodies via artificial intelligence (AI)? Is it actor Nicholas Cage? … *It's broader than either, the US Department of Defense says. Rather, it's all of us who are exposed to fake news and run the risk of getting riled up by garbage* (Article 112, emphasis added).

Throughout the articles certain types of falsified content were given more attention and concern. Indeed, the use of deepfakes for political (propaganda) purposes is identified explicitly as a "bigger worry" (Article 18) than sexual deepfakes. One article notes that while the software is "currently used mainly in porn flicks […] it may not be long, though, before fake 'political' videos are created" (Article 62); another warns that "with enough time and technological advancement, political speeches could be falsified, and 'fake news' could become a genuine threat" (Article 20). The message, though subtle, is repeated enough to be abundantly clear: the threat presented by deepfakes pornography is less serious, less genuine, and less significant than the potential political consequences of fake news and disinformation.

There is no doubt that synthetic video could undermine the electoral process and threaten national security; there is also no question that these productions could undermine our sense of a shared reality. These concerns, however, are obviously different in nature from the personal harms experienced by targets of sexual deepfakes, who face reputational harm, damage to their emotional and psychological wellbeing, and whose

content is being altered and recirculated without their consent (Bates, 2017). News coverage that treats deepfakes pornography as subordinate to fake news effectively elides the harms associated with sexual deepfakes, rendering those consequences, and the women who are affected, invisible in the public view. As mentioned previously, when the targets of sexual deepfakes were discussed, the focus was on the women whose faces were used, not the Adult performers bodies; more specifically, celebrities were often used as examples in lieu of less public-facing examples. Coverage also focused on cis-gendered women as targets of deepfakes and did not extend their discussion to transgender or non-binary individuals. These elements of coverage—public facing, cis-gendered women—suggests a tendency to prioritize certain types of harm and sends messages about whose harm matters most.

Across the coverage of deepfakes, there was recognition of a range of possible harms or problems, all rooted in the fact that deepfakes allows the production of synthetic video that is increasingly difficult to distinguish from the "real" thing. The themes of harm to the audience of these productions (e.g. creating false beliefs and undermining the political process) and harm to the targets (e.g. the women in sexual deepfakes) of these productions were both discussed in the media coverage. Given that the coverage recognized *both* audience and subjects as experiencing harm, it is somewhat surprising that in many cases the discussions suggested that one harm—the harm to targets, especially from sexual deepfakes—was subordinated to the other—the harm to the audience of fake news and political productions. This positioning is relatively subtle, and surprisingly consistent, representing not a wholesale *denial* of the problem of sexual deepfakes and the consequences for women, but instead a selective emphasis on the social and political consequences of the use of deepfakes to create and spread fake news and disinformation. Noting her impression of the response to deepfakes, Maddocks (2020) writes that there is an overwhelming focus on "political deep fakes, while their pornographic counterparts have become part of the scenery in cyberspace" (p. 2).

Surprisingly, even when articles clearly stated there was an "obvious problem" (Article 91) with deepfakes, that "obvious problem" did not focus on the *doing*, *choosing*, and *creating* of deepfakes—for pornographic or political ends. Missing entirely from this coverage is any discussion about *who* is making these deepfakes and what in our culture permits this clear transgression. Countless women have had their images doctored to make those images appear sexual in nature. The use of deepfakes for this end is part of a larger socio-cultural pattern of misogynistic behavior, and the connection between the specific problem of sexual deepfakes and the larger misogynistic social context is largely neglected in the press coverage.

Conclusion

The goal of this study was to understand how new media construct the problems associated with deepfakes. In our research, we asked the question: how does news media characterize the problems presented by deepfakes? The answer is that deepfakes is presented primarily as a problem for the audiences of the fake video productions, who are at risk of being misled by false videos produced with a new, easy to use, technology. Future and as yet hypothetical consequences of false videos, including fake news and

disinformation, election disruption, and threats to national security, consistently received more attention, both in content and context, than did the harms associated with the dominant and existing use of deepfakes to create non-consensual sexual content. The intense focus in the news coverage of deepfakes on politics and fake news promotes the need to prepare for the (inevitable) political misuse of deepfake technology. When, however, this coverage fails to distinguish the harms associated with sexual deepfakes from the potential harms of political disruption and disinformation, it demonstrates a disregard for the impact of sexual deepfakes on the lives of women, and more to the point a disregard for the social and cultural factors that underpin this use of deepfake technology. Like news media, "images might not tell the truth, but they do tell us something: what is desirable, for example, or what is worth paying attention to" (van der Nagel, 2020, p. 4). By failing to tell a more inclusive story, the articles in our sample can limit public perception and affective and institutional responses.

The continuing outright misogynistic use of deepfakes, specifically for the creation and distribution of non-consensual sexual content, ought to be an important element in any discussion of the technology. Even as the misogynistic use of deepfakes continues, the concern for the use of deepfakes to spread disinformation for politically nefarious ends grows. Our concern is that the harm caused by sexual deepfakes, and thus the harm to the women in such videos, is overshadowed by other concerns. The harm associated with sexual deepfakes cannot be subsumed under other harms: instead, they are *different* harms, which in our opinion should carry equal weight. The real danger is in the subordination of the misogynistic harms to political harms.

Notes on contributors

Chandell Gosse is a PhD Candidate in Media Studies in the Faculty of Information and Media Studies at Western University. Her research takes an interdisciplinary approach and sits most broadly at the intersection of feminism, digital culture, and anti-violence work. Find her on Twitter @ChandellEnid

Dr. Jacquelyn Burkell is an Associate Professor in the Faculty of Information and Media Studies at the Western University. Her research focuses on the social impact of technology, with a specific emphasis on privacy. In her work, she links empirical research with legal and policy outcomes.

Data availability statement

The data that support the findings of this study are openly available in fig share at https://doi.org/10.6084/m9.figshare.12098307.

ORCID

Chandell Gosse 🔟 http://orcid.org/0000-0002-9868-2796

References

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The state of deepfakes: Landscape, threats, and impact*. DeepTrace.

Bates, S. (2017). Revenge porn and mental health: A qualitative analysis of the mental health effects of revenge porn of female survivors. *Feminist Criminology*, *12*(1), 22–42. https://doi.org/10.1177/1557085116654565

Burkell, J., & Gosse, C. (2019). Nothing new here: Emphasizing the social and cultural context of deepfakes. *First Monday*, *24*(12). https://doi.org/10.5210/fm.v24i12.10287

BuzzFeedVideo. (2018, April 17). You won't believe what Obama says in this video. https://www.youtube.com/watch?v=cQ54GDm1eL0

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1752–1819.

Citron, D. K. (2014). Criminalizing revenge porn. *Wake Forest Law Review*, *49*(2), 345–391.

Cole, S. (2017, December 11). AI-assisted fake porn is here and we're all fucked. *Motherboard*. https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn

Cole, S. (2018a, January 24). We are truly fucked: Everyone is making AI-generated fake porn now. *Motherboard*. https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley

Cole, S. (2018b, February 6). Pornhub is banning AI-generated fake porn videos, says they're non-consensual. *Motherboard*. https://motherboard.vice.com/en_us/article/zmwvdw/pornhub-bans-deepfakes

Cole, S. (2018c, June 18). Deepfakes were created as a way to own women's bodies—we can't forget that. *Vice*. https://broadly.vice.com/en_us/article/nekqmd/deepfake-porn-origins-sexism-reddit-v25n2

Cole, S. (2018d, August 14). There is no tech solution to deepfakes. *Motherboard*. https://motherboard.vice.com/en_us/article/594qx5/there-is-no-tech-solution-to-deepfakes

Cox, J. (2018, February 7). Pornhub and Twitter ban deepfake videos which swap celebs' faces onto porn stars. *The Sun*. https://www.thesun.co.uk/tech/5518474/pornhub-twitter-deepfake-videos-pornstars-celebs-taylor-swift-gal-gadot-xrated/

Day, C. (2019). The future of misinformation. *Computing in Science and Engineering*, *21*(1), 108. https://doi.org/10.1109/MCSE.2018.2874117

Durant, J. (1999). Participatory technology assessment and the democratic model of the public understanding of science. *Science and Public Policy*, *26*(5), 313–319. https://doi.org/10.3152/147154399781782329

Fairclough, N. (1989). *Language and power*. Addison Wesley Longman Ltd.

Fletcher, J. (2018). Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, *70*(4), 455–471. https://doi.org/10.1353/tj.2018.0097

Galante, L., & Shaun, E. (2018, September 11). Defining Russian interference: An analysis of select 2014 to 2018 cyber enabled incidents. *Scowcroft Center for Strategy and Security*. https://www.atlanticcouncil.org/images/publications/Defining_Russian_Election_Interference_web.pdf

Johannesson, I. A. (2010). The politics of historical discourse analysis: A qualitative research method? *Discourse: Studies in the Cultural Politics of Education*, *31*(2), 251–264. https://doi.org/10.1080/01596301003679768

Kilbourne, J. (1979). *Killing us softly: Advertising's image of women*. Lazarus M and Wunderlich R (Directors). Cambridge Documentary Films.

Lather, P. (1986). Research as praxis. *Harvard Educational Review*, *56*, 257–277.

Lazar, M. M. (2007). Feminist critical discourse analysis: Articulating a feminist discourse praxis. *Critical Discourse Studies*, *4*(20), 141–164. https://doi.org/10.1080/17405900701464816

Li, Y., Chang, M. C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI generated fake facevideos by detecting eye blinking. *arXiv*. https://arxiv.org/pdf/1806.02877.pdf

Maddocks, S. (2020). 'A deepfake porn plot intended to silence me': Exploring continuities between pornographic and 'political' deep fakes. *Porn Studies*, 1–9. https://doi.org/10.1080/23268743.2020.1757499

McGlynn, C., Rackley, E., & Houghton, R. (2017). Beyond 'revenge porn': The continuum of image-based abuse. *Feminist Legal Studies*, *25*(1), 25–46. https://doi.org/10.1007/s10691-017-9343-2

Mulvey, L. (1975). Visual pleasure and narrative cinema. *Screen*, *16*(3), 6–18. https://doi.org/10.1093/screen/16.3.6

Murphy, M. (2018, September 26). Facebook reveals 'deepfake' avatars that bear uncanny resemblance to humans. *The Telegraph*. https://www.telegraph.co.uk/technology/2018/09/26/facebook-reveals-deepfake-avatars-bear-uncanny-resemblance-humans/

Olson, P. (2018, March 15). Fake-news makers, beware: This firm claims it can oust a photo with one fake pixel. *Forbes*. https://www.forbes.com/sites/parmyolson/2018/03/15/fake-news-photos-startup-oxford-facebook-google/#564f78ca5530

O'Reilly, S. (2018, February 3). Thanks to FakeApp, Nicolas Cage is about to become the biggest movie star of all time. *The Irish Times*. https://www.irishtimes.com/culture/film/nicolas-cage-is-now-being-spliced-into-every-film-ever-made-1.3376456

Powell, A., Henry, N., & Flynn, A. (2018). Image-based sexual abuse. In W. S. DeKeseredy & M. Dragiewicz (Eds.), *Routledge handbook of critical criminology* (pp. 305–315). Routledge.

Tonkiss, F. (1998). Analyzing discourse. In C. Seale (Ed.), *Researching society and culture* (pp. 245–260). Sage Publications.

van der Nagel, E. (2020). Verifying images: Deepfakes, control, and consent. *Porn Studies*, 1–6. https://doi.org/10.1080/23268743.2020.1741434

Vickery, J. R., & Everbach, T., (Eds.), (2018). *Mediating misogyny: Gender, technology, and harassment*. Palgrave Macmillan.

Wagner, T. L., & Blewer, A. (2019). "The world is no longer real": Deepfakes, gender, and the challengers of AI-altered video. *Open Information Science*, *3*(1), 32–46. https://doi.org/10.1515/opis-2019-0003