Towards a Typology of Intentionally Inaccurate Representations of Reality in Media Content

Matthew J. Davis, Per Fors

Towards a typology of intentionally inaccurate representations of reality in media content

Matthew J Davis[1][0000-0003-4159-6739] and Per Fors[1][0000-0001-6673-7935]

[1] Uppsala University, Uppsala, Sweden
matthew.davis@angstrom.uu.se
per.fors@angstrom.uu.se

Abstract. In this paper, we take a look at three concepts frequently discussed in relation to the spread of misinformation and propaganda online; fake news, deepfakes and cheapfakes. We have mainly two problems with how these three phenomena are conceptualized. First of all, while they are often discussed in relation to each other, it is often not clear what these concepts are examples of. It is sometimes argued that all of them are examples of misleading content online. This is quite a one-sided picture, as it excludes the vast amount of content online, namely when these techniques are used for memes, satire and parody, which is part of the foundation of today's online culture. Second of all, because of this conceptual confusion, much research and practice is focusing on how to prevent and detect audiovisual media content that has been tampered with, either manually or through the use of AI. This has recently led to a ban on deepfaked content on Facebook. However, we argue that this does not address problems related to the spread of misinformation. Instead of targeting the source of the problem, such initiatives merely target one of its symptoms. The main contribution of this paper is a typology of what we term Intentionally Inaccurate Representations of Reality (IIRR) in media content. In contrast to deepfakes, cheapfakes and fake news – terms with mainly negative connotations – this term emphasizes both sides; the creative and fun, and the malicious use of AI and non-AI powered editing techniques.

Keywords: Fake news, Deepfakes, Cheapfakes, propoganda, disinformation, misinformation, typology.

1 Introduction

Let us begin this paper by noting that the authors are two prolific researchers, whose tenacity for objective truth-finding in a post-truth society has won the m numerous journalistic integrity accolades across the globe. Their dedication to this cause is un-paralleled, and equal only to their endless pursuit of knowledge. Or is it? As a reader, does this make the premise of the paper more convincing? In any case, how does one critique such claims, if, for example, this paper is to be published in a reputable journal? Usually, if one has not read any of the authors' previous work, critical minds will seek some measure of verification, rather than take these claims at face value. Perhaps these

awards are listed on their respective LinkedIn profiles, and a quick internet search of the names of these awards reveals a professional looking website with social media posts and images relating to an award ceremony. Depending on one's level of cynicism and appetite for how deep the rabbit hole goes, a few layers of forgery might be enough to convince a skeptical reader that the authors are indeed trustworthy individuals, and that the academic contribution of this paper is worthy of a high impact factor. No harm done...right?

The erosion of trust in societies institutions is a present and widely discussed topic among today's scholars, not least because of the decay of appreciation for truth among the general population [1]. Whether by design or not, Latour's post-truth tools for querying the origins of scientific fact have been weaponized by a new wave of alt-right propagandists with alarmingly effective results [2]. This Liar's Dividend[1] has forced society to reckon with itself, and polarized politics to the point that we can no-longer believe what we see, unless it is first hand [3].

Let us contextualize this issue further. A unique trait among humankind, is the ability to communicate ideas across generations through written, and more recently, audiovisual (AV) media. For years, most of our information about current events came through carefully curated channels; newspaper articles, books, televised reporting, and with it, society had to exert a certain amount of trust that the information they were receiving was truthful since there were very few ways of learning about events occurring outside our personal bubbles of experience. Journalists had a measure of integrity and respect for the job, and access to printed media was controlled through education, accreditation and the fact that mass production of printed text was costly. Whilst this did not hinder the spread of propaganda, it at least limited the power to shape civil discourse to the establishment, whose strive for stability meant that unorthodox views were kept to the niches of society. There are many examples in modern times however, of unscrupulous actors manipulating various media in order to further their own agendas. From the Iraq war of 2003 [4] to the UK's most recent general elections and referenda [5], those with the keys to shape public discourse tend to wield the most political power.

With the advent of digital media and the internet, the limitations on how information is captured, shared and discussed around the world, changed. This decentralization of communicative channels has given the illusion of complete information freedom, and for the time being, it is much harder for media tycoons and political establishments to paint a subjective narrative since everyone now armed with a smartphone can share their own version of events, as they occur. Once again, the balance of power between institutional elites and the people is changing, but with it, the lines between objective truth and subjective narrative are becoming increasingly blurred. Technology, often thought of as a panacea for many of today's problems, can also exacerbate such issues.

Since the internet globalized our digital media, everyone with access to a computer was able to have a wide-reaching platform for their message, without the same level of scrutiny usually applied to traditional broadcasts. Initially, producers of audiovisual media content required some capital investment; lots of expensive hardware, human resources, and years of specialist education in order to create a masterful production of coercive storytelling. Now, everyone has the ability to create their own online "news" channel, promoting all manner of more or less informative or sensational opinion

pieces. A few years ago, one could differentiate between amateur influencers, and professionally produced content based on the special effects, or the premium feel. Yet with a little experience, access to low-cost technology has improved the user experience far beyond traditional media. Whilst this isn't necessarily a problem when it comes to pure entertainment, combine this with a business model that relies on attracting views and a general public's decreasing attention span, and you have a cultural crisis in the making, leading to the loss of truth as an exchange value [6]. Or as Postman [7] might argue, a swelling preference for entertainment over substance. In a time where the freemium (most content for free, some premium) business model effectively removes the funding for in-depth and thus costly investigations, salvaging journalistic integrity is also difficult [8]. How then, do uncritical minds differentiate between well produced content with verified "facts" from qualified professionals whose "evidence" challenges their beliefs, from the uninformed and opinionated ideologically driven views of a tech-savvy troll, whose entertaining tirades conform to and enhance the pre-existing biases of the target demographic? Addressing such a question goes beyond the scope of this paper, yet its importance should be highlighted since democracy as we know it, is consequently at risk.

2 Fake News

Donald Trump is often quoted when discussing fake news rhetoric, since he has successfully used the tactic to take advantage of this 'liar's dividend' in public political discourse. Whilst several scholars disentangle the semantic definition of fake news [9], the term is usually ascribed to unfavorable media, in order to confuse or polarize the debate in such a way as to remove any credibility from an authentic source; the general public, who are now a hyper-cynical audience, begin to question the legitimacy of the mainstream media. Far from being a novel strategy, sowing the seeds of doubt, and reaping the benefits of a skeptical public has been the go-to tactic for the rise of many successful politicians. However, it would be complacent to suggest that the most current candidates' rise to power is simply a "blip" in our collective sanity, attributed to an imbalance of knowledge and rational thinking among the general public. As argued above, there is a cultural crisis occurring on the back of a wave of new media, which certain political groups have exploited rather successfully.

The idea then, that a healthy democracy should allow deliberately deceptive speech under a banner of freedom, is a contradictory one, with roots in Popper's paradox of tolerance. In a "marketplace of ideas", the practice undermines, rather than enhances, the pursuit of truth since "people are more likely to be influenced by the first version of events that they hear, and are increasingly likely to remember falsehoods the more they are exposed to them" [10][54]. Now more than ever, wider society needs to be vaccinated against the fake news phenomenon [11], especially since the tools and techniques used by those who would subvert the democratic process are using increasingly sophisticated technologies. If we value our democracy, and strive to build a just society based on civil liberty, we must elucidate the different types of dis- and mis-information so that, despite the interests of unscrupulous actors, even the most gullible person is

equipped to treat "fake news" as an empty signifier, and reject it. This type of vaccination is categorized as an educational remedy; it relies on changing the general public from cynical, into critical, which is easier said than done. The Open Society Institute highlight in their report that nations with a higher level of education among the general public, tend to be more resilient to fake news [12]. Before we look at solutions however, let us discuss the different types of media that are designed to misrepresent reality.

3 Cheapfakes (or Shallowfakes)

We begin therefore with the manipulation of still images; a longstanding practice that became more readily accessible to the general public in the early 2000s with the rise of easy-to-use image editing software. When a picture is manipulated in this way, it is often called "shopped", with a reference to Adobe Photoshop, the most popular photo editing software. Since the 2000s, shopped images have cemented themselves as an important part of online culture, as can be found all over forums such as Reddit and 4chan, but also on social media platforms such as Facebook. Normally, original content is produced for Reddit or 4chan, and once such content becomes a meme[1], it becomes widespread (goes viral) on other more mainstream social networks.

While image manipulation is often used to entertain a certain online audience, such techniques can also be used with malicious intent. McGlynn et al. [13] discuss the use of shopping for pornographic purposes which they term "sexual photoshopping". According to DeKeseredy and Schwartz [14], such malicious intent of the software "normalize misogyny, hurtful sexuality, racism, and … seeking revenge on female ex-partners". The important point about sexual photoshopping is that it can be used for mainly two purposes; for revenge or for entertainment. The first category of revenge pornography is mainly practiced by men to humiliate or shame previous partners, though certain demographic groups are more likely to experience it [56-59]. The second type is usually shared for "entertainment purposes" on different subforums. For example, a more recent sexual photoshopping technique is called "bubbling" which, whilst proponents may argue actually covers up more of an original image, ultimately relies on quirks of the human brain to make a semi-clothed person appear naked [15]. It can then also be classified as entertainment, but still have a malicious intent, especially if the content is used to humiliate or degrade.

Other types of non-AI powered techniques are also used for this purpose. For example, in 2016 the legal counsel of the president Duerte used a pornographic video showing a woman impersonating Leila De Lima, a political opponent critical of the president's authoritarian rule in the Philippines. Similar things happened in India in 2002, when journalist Rana Aryyub was targeted with a pornographic video in order to ruin her credibility [16]. However, while this content requires some technical knowledge to produce, there are far simpler ways of ruining someone's reputation using unedited "evidence" of the person in an exposed position, namely to recontextualize an already

[1] A meme is a visual or textual expression of an idea, behaviour or style that spreads and evolves online by means of imitation and is not to be confused with "viral content", which refers to anything – a text, an image or a video – that is frequently shared online.

existing video clip or picture. Usually, a picture or video clip is posted on social media showing for example a politician saying something, with a made-up text of what the video or picture is describing. While such techniques are commonly used by propaganda outlets to mislead and manipulate, it is more often used for entertainment purposes. In such cases however, it is often clear that the context of the picture or video has been altered by the original poster.

Another low-tech technique that is often used to manipulate video clips is splicing. Splicing means that segments of a video are put together in certain way, so that a - often political - message is communicated. In 2018, for example, a false BBC report started circulating on WhatsApp, showing how a nuclear war was developing between NATO and Russia, claiming that Russia had used tactical nuclear weapons against the UK [17]. The clip also shows the royal family evacuating Buckingham Palace. The technique is also used to cheat in different online gaming competitions, most notably speedrunning[2]. Apart from splicing, there are other low-tech video editing techniques that can be used to mislead a particular audience. The most well-known is probably the "drunk Pelosi" video. This video, which was shared on Twitter also by US president Donald Trump, showed Nancy Pelosi in a debate, seemingly intoxicated. In fact, the video had been slowed down to 75% of its original speed [18].

What these techniques demonstrate, is that there are already many ways in which people attempt to control public discourse to suit their own agenda. This is not a new phenomenon; most of the general public has been primed towards them for many generations and is largely able to critique this type of content effectively. Consequently, a relatively skeptical audience exists which is somewhat immune to these types of propaganda. A problem arises however, when new technologies are developed which challenge our perceptions of reality and lead us to question whether what we see, is actually what we get.

4 Deepfakes

The term deepfake was coined after a Reddit user called Deepfakes posted several pornographic videos on a subreddit in 2017. These videos, while professionally produced, seemed to feature well-known female celebrities as pornstars. The term deepfake is a combination of the terms "deep learning" and "fake" and refers to Audio/Visual (AV) media content that has been edited using AI powered software in order to produce falsified but authentic-looking results [55]. The tools used to create deepfakes were published shortly after the first pornographic movies, and while Reddit quickly banned deepfaked pornography from their site, users of these tools found other forums to share their creations and improve their tools. The tools often use the Google Image search function, look through social media sites and replace faces in videos in a convincing manner, so long as the source material is of high enough quality. After the first machine

[2] This refers to a certain way of completing a video game or selected parts of it as quick as possible. In order to prove that the run has been completed in a certain time, the runner needs to provide video evidence.

learning process, the programs often require little to no human supervision, and the algorithms improve the process more or less autonomously [19].

The technology used in these tools is a machine learning technique called GAN, which stands for generative adversarial network, and has utility not only for AV media editing, but also in cutting edge medicine and computer science research [20], hence, an outright ban on developing such technology would be problematic for numerous reasons. In any case, the technology required to create convincing deepfakes, combined with fake voices, currently requires a large seed dataset. Fried et al. [21] demonstrate a relatively straight forward process for achieving lifelike reconstructions, though it requires a lot of time for processing and good quality media. For deepfaking the everyday person, the specificity of clear, well lit images or video, and specific spoken phrases may be quite difficult to gather and compile. For public figures who are always in the spotlight however, this can be quite easy to acquire. Indeed, with ever increasing capabilities of recording and processing technology, the barrier to entry for public use of these deepfake algorithms will continue to diminish. We therefore envisage a point in the very near future where, like much of the printed news media, one can no longer trust that visual media is representative of real life either. The impact of such a step change to our society should not be underestimated, since the institutional consequences are likely to be far reaching.

According to Zannettou et al. [22], there are a number of actors associated with the production and distribution of deepfaked content. These range from governments, political activists, criminals, malevolent individuals (paid and unpaid trolls), conspiracy theorists and social media bots, pushing the content towards certain groups of people susceptible to whatever message is being communicated. Westerlund [23] presents a similar list of the producers of this kind of content: communities of deepfake hobbyists, political players such as foreign governments, and various activists, other malevolent actors such as fraudsters, and legitimate actors such as television companies. The question of 'why' such agents feel compelled to produce deep faked content is an interesting one, if a little loaded at present. Empirically however, it is also clear that the technology itself is not entirely neutral, since the tools used to generate deepfaked content were popularized in male-dominated online communities and are mainly used for pornographic purposes.

Consequently, it is important to emphasize that, despite the growing amount of deepfaked content online, deepfakes have yet to be used for purposes of political propaganda or in a malevolent misleading way with any substantial impact. According to Deeptrace – a company that uses different techniques to detect and monitor deepfaked content – pornography, and mostly celebrity pornography, makes up approximately 96 percent of the deepfaked content [24]. While this is by no means positive, we can safely say that the average user and distributor of deepfaked content is not a fraudster or a politician with malicious intent, but for now, a tech-savvy pervert [25] or relatively harmless troll. Whilst the damaging effects of this type of content are contingent on various social stigmas surrounding pornography, this does not however diminish the risk of the technology being abused for political means once they are harder to detect. There are several well-made deepfakes which use the likeness of former US president Barack Obama

[26], UK prime minister Boris Johnson, and UK leader of the opposition Jeremy Corbyn [27], to demonstrate the alternative applications of such technology outside of pornography.

Unsurprisingly then, it is commonly assumed that "deepfakes are a major threat to society, the political system and businesses" [23]. However, while some papers mention that there are positive uses of this technology, most contributions express fear or worry about the technology itself rather than the problem of misinformation, which is described as a "growing" [21], very "real threat" [28, 29] to elections and society. Much research on the deepfake phenomenon therefore focuses on how to create technological tools for deepfake detection; indeed, most of the papers we found have this focus, and admittedly we have also considered this option.

5 A Society in Search of Solutions

Although it is a novel approach, we are not the first to consider using the blockchain to ensure the integrity and validity of video content. In a recent paper, Hasan and Salah [30] present a well thought out solution which uses the Ethereum blockchain and smart contract functionality in order to trace the history of digital content to its original creator source. For many unaware readers, the context of such a system requires some clarification. Most of the internet as we know it today is heavily centralised, and exists in a client-server relationship, with large companies such as Google, Facebook, and Amazon hosting most of the world's data in their cloud-services warehouses. As with any centralised system architecture, there are inherent advantages in terms of speed of access, security, and scalability. However, there are also many weaknesses, including maintaining anonymity and data privacy; if you can own the server through a hack or some such vulnerability, you can own the data or even replace it with your own version. Consequently, there are many projects underway to develop a decentralised system architecture which relies on blockchain technology to store the data in a more distributed and secure manner. Ethereum is one such blockchain project, though others exist with similar smart contract functionality, and each has its own marketplace of users developing dApps (decentralised applications). Indeed, there is an ongoing ideological debate occurring between these projects about just how decentralised they need to be, and it is likely that, as with most new technological standards, only a few will reach the critical mass of mainstream adoption required to survive.

Under Hasan and Salah's framework, all media content would be registered on a "decentralized, content-addressable, peer to peer file system", such as the IPFS (InterPlanetary File System). Content creators publish an original video, and subsequent edits require the permission from the original creator and are linked to the original video as records on the immutable ethereum blockchain. For the main issue of this paper however, this is problematic for several reasons. Firstly, such a system requires many content creators to switch to a new type of decentralised internet which is currently not in a user-friendly format. Secondly, anyone can still copy a published video using some basic screen capture software, and then modify and upload an 'original' copy which can then be seen as the original source of this new file. Furthermore, their "underlying

principle of solving the deepfake problem simply relies on providing undisputed traceability to the original source", which does not really address the creation or distribution of disinformation.. If the original publisher of a video uses their own captured and modified footage, how are we to know the truth of the matter? Perhaps this is not the point however, since in many ways, the solution is a more technically complex version of Twitter's "Verified" status, built for a decentralized system architecture of peer-to-peer content creation. The authors' main stated goal is to offer a workable solution for a blockchain based "Proof-of-Authenticity", and they should be commended for doing so. Indeed their solution may prove highly useful for copyright control and distribution of royalties in the near future. If the technological trend is moving towards a decentralized internet, then such a framework can succeed perhaps at least in helping users to determine whether their digital content comes from a trusted and reputable source.

In any case, the recent years' "truth decay" is certainly worrying; just as Trump uses the term "fake news" to deflect legitimate criticism away from his actions, one must now question whether we can trust what we see, since it is also possible that media content has been tampered with in some way – manually or with the help of AI. What can be done to defend against the potential negative impact of this disruptive technology on society? We do not give regulatory solutions much weight here, simply because regulation of open source technology (which also has positive uses) is impractical in any case. Criminalization of the use cases of digital technology is a separate issue which we do believe in, and has had limited efficacy against online copyright piracy, for example (though it is arguable that with the advent of cheap streaming services, the market demand for pirated material diminished). Whilst this may curb the production and spread of "revenge porn" among the general public, it does nothing to address the societal impact of misinformation, which is highly effective on an uncritical, or overly skeptical audience. Additionally, we are hesitant to suggest that regulation would inhibit governments and politicians from abusing technology via 3rd parties, since we have seen they are already extremely comfortable bending and manufacturing their version of the truth.

Some companies are attempting to combine approaches to address the issue. Facebook, who recently banned some doctored video content from their site, states that AI-powered manipulated content was rare to begin with, but that deepfaked content could potentially present a "significant challenge for our industry and society as their use increases" [31]. Banning deepfaked content certainly sounds like great news for anyone concerned about the spread of disinformation, however, Facebook is – as noted by Drew Harwell on Twitter [32] – merely targeting a certain video-editing technique and not spread of disinformation per se. For example, non-AI powered editing such as the "drunk Pelosi" video mentioned above can still be shared on the site, while "dank memes" such as the deepfaked Nicholas Cage videos can potentially be removed. Furthermore, as most deepfaked content can already be classified as pornography, Facebook already has a policy against the vast majority of malicious deepfaked content.

Given the strong academic and civil discourse expressing worry and fear for the technology itself and not mainly in which way it is used, it is time to put the technology in a larger context which highlights the fact that media content is manipulated for many different reasons, and that manipulation of media content is a big part of online culture.

The same thing goes for satire and parody articles which are sometimes casually described as fake news, although they do not seek to mislead but to entertain (see e.g. [9, 33]). More importantly, while being used for many different purposes and with very different motives, fake news, cheap fakes and deepfakes have more in common than the research community has yet realised. The aspects mentioned above are highlighted in our typology below, where we collectively refer to different types of fakes as Intentionally Inaccurate Representations of Reality (IIRR) in media content (Table 1).

Table 1. A typology of IIRR in media content

Type	Media format	Purpose	Distributed through	Uses AI	Potential of malicious use	Also known as
Fake news	Textual, AV	Mislead	Traditional media, social media, blogs, word of mouth (wom)	No	High	"Fake news"
Polarized content/biased content/misreporting/selective reporting	Textual, AV	Mislead, create opinion	Traditional media, social media, blogs, wom	No	Medium	"Fake news"
Parody	Textual, AV	Humour, memes	Traditional media, social media, blogs	Yes /No	Low	"Fake news"
Satire	Textual, AV	Humour, memes, propaganda	Traditional media, social media, blogs	Yes /No	Low	"Fake news"
Citizen journalism	Textual, AV	Mislead, create opinion	Social media, blogs, wom	No	High	"Fake news"
Clickbait	Textual, AV	Mislead, generate clicks	Social media	No	Medium	"Fake news"/"Click bait"
Conspiracy theory/pseudoscience	Textual, AV	Mislead, generate opinion, generate clicks	Social media, blogs, wom	No	High	"Fake news"
Image enhancement	Image	Improve quality of images	Traditional media, social media	Yes /No	Low	N/A
Automated photo editing	Image	Improve quality of images	Traditional media, social media	Yes	Low	N/A
Splicing	AV	Humour, memes, mislead	Social media, blogs	No	High	"Cheapfake"/"Shallowfake"

Type	Media format	Purpose	Distributed through	Uses AI	Potential of malicious use	Also known as
Photo editing (Photoshopping)	Image	Humour, sexual photoshopping, memes, generate opinion, create opinion, generate clicks	Traditional media, social media, blogs	No	N/A	"Cheapfake"/"Shallowfake"/"Shopped"
Face Swapping/Face-morphing/ Full-body puppetry	Video, Image	Humour, memes, propaganda, sexual photoshopping, generate opinion	Social media	Yes	Low	"Deepfake"
Lip syncing and voice synthesis	AV	Humour, memes, propaganda	Social media, forums, blogs	Yes	N/A	"Deepfake"
Re-contextualization	AV	Humour, memes, propaganda	Social media, blogs	No	High	"Cheapfake" / "Shallowfake"
Text-to-speech/voice-swap	Audio	Humour, mislead, propaganda	Social media, blogs	Yes /No	Medium	N/A
Astroturfing	Textual, AV	Mislead, propaganda, generate opinion	Social media, blogs, traditional media	Yes /No	High	Covert lobbying, fake grassroots

The IIRR typology has been put together through a literature review of papers that have tried to in some way categorize either deepfaked or cheapfaked content (e.g. [16, 34]) or fake news (e.g. [31–34]). The level of "potential of malicious use" has been evaluated based on our own perception of the type of IIRR content combined with other researchers' perceptions. The typology itself is a work-in-progress and therefore we invite other researchers to continue this work in future research contributions in order to refine and perfect the typology. Such a typology will be useful in categorizing and identifying misuses of future technological improvements, and perhaps for building defensive strategies, whether through regulatory, educational, or technological means.

It is important to note that despite the fact that technology itself is not neutral, its potential is not entirely negative. As Silbey and Hartzog [8] suggest, perhaps the deepfake phenomenon is precisely what society needs to properly reflect and deal with the breakdown of trust in political institutions. By challenging the very notion of 'seeing is believing' innate to our concept of trust, established power structures will no longer be able to capitalize on an uncritical audience. That being said, it will be difficult to muster

the political will to modify these institutions sufficiently if they continue to benefit from a disenfranchised general public. We suggest more focus on building genuine grass-roots organizations which challenge the meta-narrative of deepfakes through improved public education and building a groundswell of political support to tackle the systemic issues and causes of fake news efficacy. We repeat, in this context, deep fakes are not the problem; the socio-economic models which lead to truth decay in the political sphere are.

6 Conclusions

In this paper, we have argued that the academic and civil discourse around fake news, cheapfakes and deepfakes is potentially harmful for our perception of the problems that these techniques give rise to. What we have observed is that a majority of the authors of papers related to these concepts implicitly or explicitly express concern for the development of these technologies per se, rather than the overall phenomenon of misinformation.

For example, Westerlund [34], in a review of deepfake research, argues: "The reviewed news articles suggest that there are four ways to combat deepfakes: legislation and regulation, corporate policies and voluntary action, education and training, and anti-deepfake technology". Concerning fake news, many articles promote so-called fake news detectors or identifiers (e.g. [37, 39–45]). For deepfake research, it is more of the same. In most of the reviewed papers, the authors are discussing how to either detect or prevent deepfakes (e.g. [46–51]). Thus, while many researchers argue that we need a healthy mix of regulatory support, corporate policies, education and training and anti-deepfake technologies, most of the research on deepfakes focuses on anti-deepfake technologies. Consequently, we note a strong urge within the field to attempt to solve a societal problem, i.e. malevolent actors spreading disinformation and non-consensual pornography, with purely technological means [52].

We have thus produced the beginnings of a typology of intentionally inaccurate representations of reality in media content (IIRR), which aims to help identify and inform consumers of media content, the ways in which information can mislead, both for entertainment and more nefarious purposes. The next iteration of this paper will aim to incorporate a more systemic approach to this analysis, alongside comments for improvement from our scholarly peers.

Additionally, this paper argues that although the technology itself is not neutral, the existence of these technologies themselves is not the problem. When it comes to deepfakes, Pandora's box has already been opened, and there is no way to "uninvent the bomb" [53]. It is reasonable to view the distribution of deepfaked content as yet another way of manipulation of the truth for malicious schemes, but – as with more traditional forms of textual manipulation – also for satire, humour and memes which is the foundation for much of today's online culture. Deepfaked content is just a new and more technologically complex expression of that culture. Hence, we agree with Gutiérrez-Martín et al's assertion that "the problem of misinformation is not solved by attacking the symptoms with a series of tips and checklists for consumers to learn how to detect

fake news but, rather, by studying its underlying causes, one of which is excessive monetization and the diminishing value of truth in new virtual environments" [6]. Undoubtedly, over time the general public will once again become familiar and resilient to modern techniques, but how much damage will be done to the fabric of our society in the interim, if there is a complete breakdown of trust in our institutions?

References

1. Chesney, R., Citron, D.: Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Aff. 98, 147 (2019)
2. Kofman, A., Kofman, E.: Bruno Latour, the post-truth philosopher, mounts a defense of science. New York Times. 6 (2018)
3. Eichensehr, K.: Don't Believe It If You See It: Deep Fakes and Distrust. Jotwell J. Things We Like. 1, 1 (2018)
4. Miller, D.: Tell me lies: propaganda and media distortion in the attack on Iraq. Pluto Press (2004)
5. Department of Digital, Culture, Media, and Sport: Disinformation and 'fake news': Final Report: Eighth Report of Session 2017–19. 1–109 (2019)
6. Gutiérrez-Martín, A., Torrego-González, A., Vicente-Mariño, M.: Media education with the monetization of YouTube: The loss of truth as an exchange value. Cult. y Educ. 31, 267–295 (2019). https://doi.org/10.1080/11356405.2019.1597443
7. Postman, N.: Amusing ourselves to death: Public discourse in the age of show business. Penguin (2006)
8. Silbey, J., Hartzog, W.: The Upside Of Deep Fakes (2019)
9. Molina, M.D., Sundar, S.S., Le, T., Lee, D.: "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. Am. Behav. Sci. (2019). https://doi.org/10.1177/0002764219878224
10. Franks, M.A., Waldman, A.R.I.E., Chesney, B., Dogan, S., Hartzog, W., Jurecic, Q., Kadri, T., Silbey, J., Sylvain, O., Wittes, B.: Sex, Lies , And Videotape: Deep Fakes And Free Speech Delusions (2019)
11. Roozenbeek, J., Van Der Linden, S.: The fake news game: actively inoculating against the risk of misinformation. (2018). https://doi.org/10.1080/13669877.2018.1443491
12. Lessenski, M.: Resilience To 'Post-Truth' and Its Predictors in the New Media Literacy Index 2018 *. Open Soc. Inst. Raporu. Mart, (2018)
13. McGlynn, C., Rackley, E., Houghton, R.: Beyond 'revenge porn': The continuum of image-based sexual abuse. Fem. Leg. Stud. 25, 25–46 (2017)
14. DeKeseredy, W.S., Schwartz, M.D.: Thinking sociologically about image-based sexual abuse: The contribution of male peer support theory. Sex. Media, Soc. 2, 2374623816684692 (2016)
15. Urban Dictionary: Bubbling, defined by user "4Red," https://www.urbandictionary.com/define.php?term=Bubbling
16. Paris, B., Donovan, J.: Deepfakes and Cheap Fakes. (2019)
17. Coulter, M.: BBC issues warning after fake news clips claiming NATO and Russia at war spread through Africa and Asia, (2018)
18. Sadiq, M.: Real v fake: debunking the "drunk" Nancy Pelosi footage - video, (2019)
19. Maras, M.-H., Alexandrou, A.: Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. Int. J. Evid. Proof. 23, 255–262 (2019). https://doi.org/10.1177/1365712718807226

20. Beridze, I., Butcher, J.: When seeing is no longer believing. Nat. Mach. Intell. 1, 332–334 (2019). https://doi.org/10.1038/s42256-019-0085-5
21. Fried, O., Tewari, A., Zollhöfer, M., Finkelstein, A., Shechtman, E., Goldman, D.B., Genova, K., Jin, Z., Theobalt, C., Agrawala, M.: Text-based Editing of Talking-head Video. 38, (2019)
22. Zannettou, S., Sirivianos, M., Blackburn, J., Kourtellis, N.: The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. J. Data Inf. Qual. 11, 1–37 (2019)
23. Westerlund, M.: The Emergence of Deepfake Technology: A Review. Technol. Innov. Manag. Rev. 9, (2019)
24. Ajder, H., Patrini, G., Cavalli, F., Cullen, L.: Report 2019: The State Of Deepfakes. (2019)
25. Öhman, C.: Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography. Ethics Inf. Technol. 1–8 (2019)
26. BuzzFeed and Jordan Peele: You Won't Believe What Obama Says In This Video! ;). BuzzFeedVideo (2018)
27. Future Advocacy: Deepfakes, https://futureadvocacy.com/deepfakes/
28. Maddox, T.: Here are the biggest IoT security threats facing the enterprise in 2017. Consult. https//www. techrepublic. com/article/here-are-the-biggest-iot-security-threats-facing-the-enterprise-in-2017. (2016)
29. Dack, S.: Deep Fakes, Fake News, and What Comes Next, (2019)
30. Hasan, H.R., Salah, K.: Combating Deepfake Videos Using Blockchain and Smart Contracts. IEEE Access. 7, 41596–41606 (2019). https://doi.org/10.1109/ACCESS.2019.2905689
31. Chappell, B.: Facebook Issues New Rules On Deepfake Videos, Targeting Misinformation, (2020)
32. Twitter: Drew Harwell. (2020). [online] https://twitter.com/drewharwell/status/1214394479026855936
33. Farkas, J., Schou, J.: Javnost-The Public Journal of the European Institute for Communication and Culture Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. (2018). https://doi.org/10.1080/13183222.2018.1463047
34. Westerlund, M.: The Emergence of Deepfake Technology: A Review. (2019)
35. Bovet, A., Makse, H.A.: Influence of fake news in Twitter during the 2016 US presidential election. Nat. Commun. 10, 1–14 (2019)
36. Wu, L., Liu, H.: Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: Proceedings of the eleventh ACM international conference on Web Search and Data Mining. pp. 637–645 (2018)
37. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: Methods for finding fake news. Proc. Assoc. Inf. Sci. Technol. 52, 1–4 (2015)
38. Waszak, P.M., Kasprzycka-Waszak, W., Kubanek, A.: The spread of medical fake news in social media – The pilot quantitative study. Heal. Policy Technol. 7, 115–118 (2018). https://doi.org/10.1016/j.hlpt.2018.03.002
39. Wang, W.Y.: " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv Prepr. arXiv1705.00648. (2017)
40. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. ACM SIGKDD Explor. Newsl. 19, 22–36 (2017)
41. Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S., de Alfaro, L.: Some like it hoax: Automated fake news detection in social networks. arXiv Prepr. arXiv1704.07506. (2017)

42. Ruchansky, N., Seo, S., Liu, Y.: Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 797–806 (2017)
43. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. arXiv Prepr. arXiv1708.07104. (2017)
44. Long, Y., Lu, Q., Xiang, R., Li, M., Huang, C.-R.: Fake news detection through multi-perspective speaker profiles. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 252–256 (2017)
45. Granik, M., Mesyura, V.: Fake news detection using naive Bayes classifier. In: 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). pp. 900–903. IEEE (2017)
46. Li, Y., Lyu, S.: Exposing deepfake videos by detecting face warping artifacts. arXiv Prepr. arXiv1811.00656. (2018)
47. Güera, D., Delp, E.J.: Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 1–6. IEEE (2018)
48. Koopman, M., Rodriguez, A.M., Geradts, Z.: Detection of deepfake video manipulation. In: The 20th Irish Machine Vision and Image Processing Conference (IMVIP). pp. 133–136 (2018)
49. Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., Li, H.: Protecting world leaders against deep fakes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 38–45 (2019)
50. Matern, F., Riess, C., Stamminger, M.: Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). pp. 83–92. IEEE (2019)
51. Korshunov, P., Marcel, S.: Vulnerability assessment and detection of Deepfake videos. In: The 12th IAPR International Conference on Biometrics (ICB). pp. 1–6 (2019)
52. Fors, P.: Problematizing Sustainable ICT, (2019)
53. MacKenzie, D., Wajcman, J.: The social shaping of technology: How the refrigerator got its hum. Milton Keynes–Philadelphia. Milton Keynes (1985)
54. Fazio, L. et al.: Knowledge Does Not Protect Against Illusory Truth. Journal of Experimental Psychology: General. Vol 144, No.5, pp. 993-1002. (2015).
55. Merriam-Webster.: Words We're Watching: 'Deep Fake'. Merriam-Webster [online]. Accessed 24.07.2020. Available here: https://www.merriam-webster.com/words-at-play/deep-fake-slang-definition-examples
56. Wolak, J., Finkelhor, D.: Sextortion: Findings from a survey of 1,631 victims. Crimes Against Children Research Center, University of New Hampshire. (2016).
57. Lenhart, A., Ybarra, M. Price-Feeney, M. Nonconsensual Image Sharing: One in 25 Americans has been a Victim of "Revenge Porn". Data & Society Research Institute, CiPHR, Data Memo. (2016).
58. O'Connor et al. Cyberbullying, revenge porn and the mid-sized university: Victim characteristics, prevalence and students' knowledge of university policy and reporting procedures. John Wiley & Sons Ltd, Wiley Online Library. Higher Education Quarterly, 72, pp. 344-359. (2018)
59. Citron, D., & Franks, M. Criminalizing revenge porn. Wake Forest Law Review, 49(2), 345-392. (2014).