Who Are the Targets of Deepfakes?

Evidence From Flagged Videos on TikTok and YouTube

Carl Öhman

Uppsala University, Sweden

✉ carl.ohman@statsvet.uu.se

Abstract

This study draws on a dataset of 1394 flagged deepfakes from YouTube and TikTok annotated by the race, gender, profession and nationality of the 'targeted' individuals. The findings indicate that, by stark contrast to pornographic sites, mainstream deepfakes overwhelmingly feature individuals who are white, male, and—when in politics—authoritarian. A complementary qualitative analysis suggests that very few videos have a political edge, even when featuring politicians. Instead, it is proposed that the chief function of mainstream deepfakes is to be seen as a *carnivalesque* (apolitical) profanation of the faces of power. These findings challenge previous studies on the demographics of deepfake targets and broaden the scope of the current literature by shedding further light on how AI-assisted video doctoring is used beyond porn and disinformation.

1. Introduction

The academic as well as popular literature on deepfakes has mushroomed in recent years, with two foci being particularly salient (see Gosse & Burkell, 2020). The first is pornography, which is often argued to be the primary context in which synthetic media technologies are employed. A 2021 report by Deeptrace (Ajder et al., 2019) claimed that as much as 96% of deepfakes were pornographic, and that 100% featured faces of female celebrities, primarily from the United States and South Korea. A subsequent report from Home Security Heroes (2023) similarly found that 98% of existing videos were pornographic, of which 99% featured female targets. Even though both reports rely on a questionable methodological design, these numbers suggest that the deepfake phenomenon is perhaps primarily a pornographic one. This has led feminist scholars as well as digital ethicists to call attention to the various ways in which deepfakes threaten women, both as individuals and as a collective (Harris, 2019; Maddocks, 2020; Öhman, 2019), and law scholars have begun to explore the various legal options to combat the phenomenon (Harris, 2019; Mashinini & Africa, 2020; Perot & Mostert, 2020).

The second focus in the literature has been political disinformation (Citron & Chesney, 2019; Diakopoulos & Johnson, 2021). Here, the problem has largely been framed in epistemic terms, and sometimes with straight-out apocalyptic undertones. Indeed, words like "epistemic apocalypse"

(Habgood-Coote, 2023), "infocalypse" (Schick, 2020a), or "epistemic maelstrom" (Rini, 2008, 8) signal that the advent of synthetic media is about to destroy the last shards of ground truth in an already polarized media landscape. Given these warnings, the vast majority of literature on deepfakes, particularly within the technical sciences, has come to focus on means of detection, i.e., how to tell reality from fake (see literature reviews by Godulla et al., 2021; Westerlund, 2019). Yet, the apocalyptic narratives have lately received pushback or been nuanced by a series of scholars with more modest verdicts (Habgood-Coote, 2023; Kalpokas & Kalpokiene, 2022; Kerner & Risse, 2021; Nieweglowska et al., 2023). Synthetic media content can already be indistinguishable from reality, yet to date, there is no known case where a significant share of the public has fallen for a doctored media artefact for an extended period of time. Even the infamous synthetic images of Donald Trump's arrest, commonly used to illustrate the dangers of deepfakes, were only spread for a (relatively) brief moment before being removed from mainstream platforms. As suggested by Nieweglowska et al. (2023, 1), deepfakes are not necessarily shared as a means of persuasion, but rather to reinforce one's "identity and social position." Thus, as Kalpokas & Kalpokiene (2022, 48) note, "much of the current debate about deepfakes is ignited more by panic than by fact and reasonable prediction."

Naturally, the above are praiseworthy pursuits, and even some of the apocalyptic warnings deserve to be taken seriously, at least in the long-term. Yet, an excessive focus on hypothetical doomsday scenarios poses the risk of overshadowing what is already going on. The significance of synthetic media lies not only in what it can potentially do in the future, but in how the technology is already employed—in its current role in society. And in today's media landscape, most people come into contact with deepfakes through mainstream platforms like YouTube and TikTok (see data on traffic below) which have stricter rules on both mis/dis-information and nudity than do alternative platforms. For example, both YouTube and TikTok require doctored content to be clearly disclosed as such, regularly remove deceptive videos that may pose political harm, and outright bans pornographic content and nudity.[1] Hence, whatever deepfakes are used for on the big video-sharing platforms, it falls beyond the two current foci of academic literature. In view of this lacuna, the goal of the present article is to provide a baseline knowledge about whom is targeted in deepfakes on the major video-sharing platforms and what cultural function these videos serve since. As such, the present study sets out to answer the following research questions:

> RQ1: Who is featured in deepfakes shared on mainstream platforms?
> RQ2: What is the cultural function of deepfakes shared on mainstream platforms?

These questions shed light on how (non-pornographic) synthetic media are actually shared and consumed within the mainstream. They allow for a more informed and nuanced debate, not least in the context of politics where focus has been placed almost exclusively on disinformation. For instance, what is the goal of doctoring videos of political figures, if not to deceive? Are liberal politicians more likely to be targeted than conservative ones? Are women more commonly targeted than men? These types of questions become possible only in light of a thorough base-line knowledge about the deepfakes shared on social media.

To answer the research questions, the study draws on a set of 1394 videos scraped using the search term "deepfake" on TikTok and YouTube, which is manually annotated by the gender, occupation, race and nationality of the individuals targeted in the videos. The findings suggest that, by stark contrast to e.g., pornographic sites, deepfakes on the mainstream services overwhelmingly use the faces of white men in the entertainment industry or in politics. They appear to be used, not so much to formulate a political or social critique, but as a means to augment artistic imagination, showing off technological

---

[1] For a full list of restrictions, see https://www.youtube.com/howyoutubeworks/our-commitments/fighting-misinformation/ and https://www.tiktok.com/community-guidelines/en/integrity-authenticity/

savvy, or what may be interpreted as a form of "carnivalesque" profanation. In other words, much of the current significance of synthetic media lies beyond disinformation and epistemic collapse—not in its ability to deceive but in its ability to enhance and animate users' imagination.

2. Methodology

The methodological procedure of this study has three components: 1) data collection, including manual annotation of the data, 2) a descriptive statistical analysis of the demographic composition of the targeted individuals, and 3) a qualitative analysis of a selected sample of the data.

2.1 Data

To get to the videos that ordinary internet users are most likely to encounter in everyday life, data were collected from the two largest venues for sharing (non-deceptive) deepfakes—YouTube and TikTok. The two platforms provide similar services, but also have notable differences.

YouTube started as a web-based video-sharing platform (notably to share a video of the accidental exposure of Whitney Huston's nipple during the 2004 Superbowl half-time show (see Sheffield, 2020), but was acquired by Alphabet in 2006. As of 2023, it has become world's second largest social media platform, with 2.86 billion monthly active users world-wide, which accounts for about half of the world's internet users. 400 hours of video is uploaded to the site every minute, making it by far the largest depository of video material in history. Due to YouTube's policy on nudity and sexually explicit content, the site has not been used to share deepfake pornography. According to its misinformation policy, it also prohibits content that has been "technically manipulated or doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm" (Youtube, n.d.). As an example of what may cause such "egregious harm," the policy cites "a video that has been technically manipulated to make it appear that a government official is dead." The practical implication of the policy has largely been interpreted as a requirement to flag the synthetic nature of deepfakes in the caption, though it has not led to a complete vanishing of deepfakes that do not declare themselves as such.

Compared to many other social media, YouTube provides a very versatile platform. It is used for such diverse purposes as political communication, learning, marketing, identify formation and sheer leisure (see Soukup, 2014). The user base includes everything from Hollywood production conglomerates to political parties and private individuals. Though it has a strong and accurate recommendation algorithm, the platform's utility is arguably too broad to afford the kind of handheld "binging" commonly associated with TikTok (see Schellewald, 2023 below).

Unlike YouTube, TikTok emerged largely as a mobile app rather than a web page. And unlike most global social media platforms, it emerged from China. The platform attracts roughly half the number of users that YouTube has (1 billion 2023), but is highly popular among young people, and has maintained a rapid growth since its start. TikTok has a stricter policy on doctored content than does YouTube, possibly as a result of pressure from American policymakers who have expressed serious doubt in the company's credibility and democratic legitimacy.[2] For example, the terms of service demand that users clearly flag manipulated content, but also prohibits deepfakes that use faces of private individuals or minors, even when this is clearly stated by the channel (TikTok, 2023). However, like YouTube, TikTok appears to enforce this policy mainly on deepfakes that are deemed to pose political threats—there are plenty of doctored videos that are not marked as such but are arguably too unrealistic to pose any serious threat of deception.

While providing features that already exist on other platforms (especially YouTube), TikTok stands out from its competitors in at least two respects. The first is the aggressive personalization of its algorithm (Bhandari & Bimo, 2022; Schellewald, 2023). Almost from the very moment users first open the app,

---

[2] At the time of writing at least the state of Montana has issued a complete ban of the app, but a nationwide closure is also looming.

they encounter content that is closely tailored to their interests in a way that other platforms have not been able to keep up with. The second feature is its situated usage. Whereas platforms like YouTube are designed with the active consumption of the content in focus, TikTok has managed to become an integrated part of many users' daily routine. Rather than each video providing an "experience," the app provides an ongoing buzz that keeps users culturally up-to-date and releases them from boredom while at the same time demanding very little in return (see (Bhandari & Bimo, 2022, 9). "It's almost like a bridge between living room and bed" as one of Schellewald's (2023, 1573) interviewees put it.

Despite the above differences, the platforms are comparable for the purposes of this study in that they are both mainstream video-sharing services whose main technological function is to provide users with content they are likely to enjoy.

Data were scraped from both services using the custom data scraping tool Apify, specifically the YouTube Scraper and TikTok Scraper features. For YouTube, the scraper collected all hits for the search-term "deepfakes," which resulted in a total of 465 items. This number does not appear to capture the totality of videos that mention the term "deepfake" in the description, which is likely due to the fact that YouTube's search algorithm (like Google's) seldom presents an exhaustive list of hits, but provides different results depending geographical region (in this case the United States). There is, however, no reason to believe that this introduces any systematic bias to the results. At least, there is no commonly accepted way of circumventing the platform's own search function.[3] For TikTok, the scraper retrieved all videos tagged with "deepfake" resulting in a total of 949 hits. This list appears to be more comprehensive than that from YouTube but may potentially suffer from similar limitations regarding scope.[4]

Upon retrieval, the datasets were manually annotated to identify the individuals whose faces or voices had been replicated in the videos (i.e., the "targets") and their demographic characteristics. This approach diverges from the aforementioned report by Deeptrace (Ajder et al., 2019) which attempted to automate the identification process by parsing through the video URLs using the python library SpaCy which includes a name recognition feature (notably, the authors focus on pornographic content only).[5] While efficient, such an approach is not very accurate. Indeed, many videos in the present datasets were not tagged with any names, despite using the faces of several famous individuals. Some videos tagged names of individuals that were not featured (supposedly to get a wider reach) and others tagged the names of fictional characters rather than the actors who play them (e.g., "Aragorn as Harry Potter"). For this reason, the material was annotated manually. Table 1 presents a list of variables and values by which the data were annotated.

---

[3] One possibility is going through Google search, but that merely switches the bias of one search engine for another's.

[4] It may be noted that the two searches were performed 5 months apart (TikTok in March 2023, YouTube in August), although this should not introduce any significant bias since the focus here is not to compare the platforms, but to get a representative sample on what kind of deepfakes users may encounter in mainstream social media.

[5] The Home Security Heroes report arguably also uses an automated process, though unlike Deeptrace they reveal nothing about their methodology.

Table 1. Variables and values for annotation

Variables	Values
Type of video	Commentary/Not Commentary/Not Related to Deepfakes
Type of target	Cartoon/Natural person
Type of doctoring	Voice/Video
Gender	Male/Female/Other
Race	White/Black/Latino/Asian/Middle Eastern
Nationality	Country Name
Name (if identifiable)	Name
Occupational sector (if identifiable)	Entertainment/Politics/Other
Political affiliation, if politician	Liberal, Conservative, Authoritarian/Populist[6]

This methodology relies heavily on the coder's personal ability to recognize the featured individuals (in cases where the text did does disclose), yet even in cases where an individual natural person was impossible to identify, demographic qualities such as race and gender could still be annotated. The annotation was performed by a single person. The full datasets and a detailed coding protocol including definitions is provided at the following GitHub repository: https://github.com/carlohman/Code-for-deepfake-analysis.git

2.2 Analytical procedure

The analytical procedure consisted of two steps. In the first step, a series of descriptive statistical measurements were made to illustrate the demographic breakdown of the targets. A multivariate regression analysis was also performed using a T-test to measure the effects of the various features of the videos and targets on the number of views. All statistical analysis were performed in R. The code can be accessed on the aforementioned GitHub depository.

In addition to the mere statistical description of the content, a sub-sample of 100 videos per site were selected for an interpretive qualitative analysis, focusing on the surrounding context and the relationship between the individual targeted and the original source. The process was guided by the following (interpretive) questions:

- What appears to be the intended effect/meaning of the video (e.g., comedy, humiliation, celebration, etc.)?
- What is the relationship between the targeted individual and the individual/character onto whom they are superimposed (e.g., opposite gender, enemies, friends, colleagues)?
- How does the video appear to be received by the audience (e.g., reactions in the comments)?
- What is the relationship between the video in question and other clips using synthetic media (e.g., following a trend, responding to previous videos)?

The goal of this study, as mentioned in the previous section, is to establish a baseline knowledge about non-deceptive deepfakes—quantitatively and qualitatively. As such, this second step is not meant to present a complete ethnographic review of the material, but is intended as a complement to the statistical

---

[6] Assessment of whether heads of state and members of the ruling party are to be listed as authoritarian/populist is based on The Economist Intelligence Units democracy index (2024). Cut-off is made at 6 and lower. This covers the vast majority of cases. In the six cases where the target is from a democratic country, yet not clearly call within the liberal nor conservative category, Levitsky and Ziblatt's (2019) four indicators of anti-democratic behavior is used as the basis for assessment.

analysis, and as a basis for further qualitative analysis along the lines of e.g., (Allison, 2021; Ayers, 2021; Glick, 2023; Holliday, 2021).

3. Findings

Overall, the findings show that, by stark contrast to pornographic deepfake sites, the targets of deepfakes on mainstream platforms are overwhelmingly male (YouTube 89%; TikTok 72%) and white (YouTube 87%; TikTok 61.5%). They primarily belong to two sectors—entertainment and politics, and within the latter category, the biggest share are authoritarian/populist. While the videos target demographics that are typically associated with social and political privilege, they are neither to be interpreted as satire nor mere humiliation/degradation. Rather, they appear to hold three key cultural functions: (1) Augmentation of artistic imagination, (2) showing off technological savvy, and (3) mere amusement, or what may be referred to as a form of *carnivalesque profanation.* Below, these findings are further detailed alongside possible explanations of the emerging patterns.

3.1 General numerical overview

Not all hits proved relevant for the demographic analysis. For example, out of the initial 949 TikTok hits, 227 were categorized either as news reporting/commentary, did not involve any sign of doctoring (such as parodies of deepfakes), or were deepfakes of cartoon characters. The remaining 722 used faces or voices of natural persons. On YouTube, half of the 465 videos were commentary and/or tutorials on how to create deepfakes, and 53 were cartoons, leaving 137 videos featuring natural persons. While the commentary/cartoons may not be relevant for the demographic breakdown, it is notable that almost half of the deepfake related content that users may encounter on YouTube is meta-content (videos commenting on the phenomenon). Or put differently, the hype seems to overshadow the actual visibility of doctored content, at least on YouTube.

The videos were posted by 356 different channels, 84 from YouTube and 272 from TikTok, giving an average of roughly 2.2 videos per channel. In total, the videos had a little over 361 million views on YouTube, and 668.7 million on TikTok, i.e., a total of a little over one billion. The view count is very unevenly distributed across the videos (and channels). The top 22 videos on TikTok account for almost a third of the views, and on YouTube, the top 7 channels stand for almost a third of the views.

To put these numbers in context, the Home Security Hero report (2023) found that, in 2023, the total number of views on the top 10 dedicated deepfake pornography sites was a little over 303 million. Even if it is difficult to make an exact comparison, these numbers indicate that the consumption of deepfakes on mainstream sites is comparable in size to, or even larger than, the pornographic side of the phenomenon. Given that the current sample probably only captures a fraction of the actual deepfakes on the platforms, whereas the data from the pornographic sites covers all videos, the difference is presumably even larger, possibly by orders of magnitude

3.2 Male/female ratio

The first thing that stands out in the demographic breakdown of the data is the overwhelming dominance of male targets (see Figure 1). On YouTube, only about 1 in 10 of the targets is a woman. The ratio is a little more balanced on TikTok, where 72% of the targets are male, but this is almost exclusively due to a single dance trend which uses the face of an unidentified Asian woman. If excluding this trend, only 9% of the videos use the face of a female target. No individuals within either platform fell within the "other" category of gender.

Proportions of sexes, TikTok deepfake targets
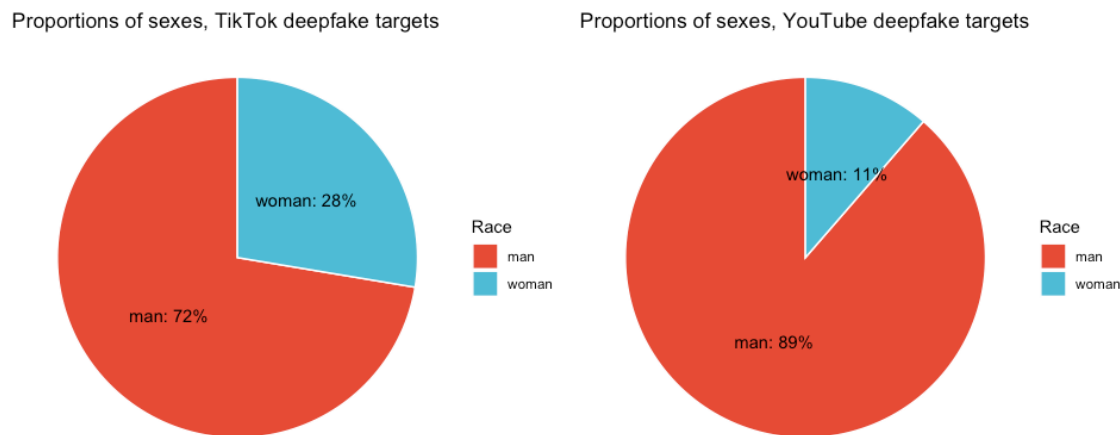
Proportions of sexes, YouTube deepfake targets

Figure 1: Gender distribution among deepfake targets on TikTok and YouTube

The numbers appear to support previous claims that, at least within YouTube video artist communities, mainstream deepfakes largely target white male "stars" (see Holliday, 2021). There are a number of possible hypotheses that may explain this dominance. The first, and most obvious one, is that deepfakes, when not pornographic, generally target famous and/or powerful individuals, and most people in power are men. Had more women been in positions of political power, there would possibly have been more deepfakes using their faces. While intuitive, this explanation has weaknesses. Only 6 videos on TikTok, and none on YouTube, feature female politicians, which is far below the female representation in politics in most parts of the world. Moreover, the women that are targeted are primarily Hollywood actresses and singers (see below) not politicians, and in the data, this category too is dominated by middle-aged white men. While it is true that, overall, men dominate in Hollywood casting, middle-aged white men can hardly be said to be the "it-boys" of the day in terms of media visibility. Indeed, in terms of internet culture, someone like Sylvester Stallone has little on actresses like Zendaya or Florence Pugh. The underrepresentation of women in positions of power hence poses a relevant, yet insufficient explanation.

Another possible hypothesis is that superimposing male faces is less stigmatized than using female ones, due to the latter being closely associated with pornography and harassment. Content creators may feel less guilt when using the face of, e.g., Stallone, than Jennifer Lawrence, because of the social context of power asymmetry between the sexes. While possible, such a hypothesis needs to be confirmed by further qualitative analysis. Are those creating deepfakes on TikTok and YouTube really the kind of people who feel concerned about such matters? Only further investigations can tell, but given the fact that at least 26% of deepfake porn consumers feel guilty about it, and that 73% of them would report to the authorities were someone close to them to fall victim to such content, there appears to be at least a general awareness of the stigma and possible harm even among users (see Home Security Hero, 2023).

Even notwithstanding these explanations, the male dominance is remarkable, especially in contrast to the pornographic sites, which nearly exclusively feature women. Future research should investigate this pattern more closely. Why are men so over represented in deepfakes on mainstream platforms when other sites have it the other way around? How do creators explain this? And what does it say about the cultural function of the phenomenon?

3.3 Race and nationality

The second thing that stands out in the data is the dominance of white targets. 87% of the YouTube targets are white, and the rest Asian, Black or "other." On TikTok, 61.5% are white, 23.7% Asian, 10.4% Black and the remaining 4.4% were "other races" (see Figure 2). However, almost all Asian targets on TikTok

can be attributed to two so-called *trends* (a format repeated by multiple private users). If these are subtracted from the data, the diagram turns overwhelmingly in favor of whites. While the distribution is not too far from the racial composition of the United States, the two platforms are used virtually all around the world and so one may have expected a more even distribution.
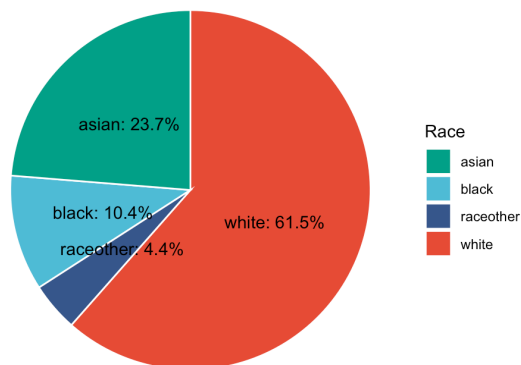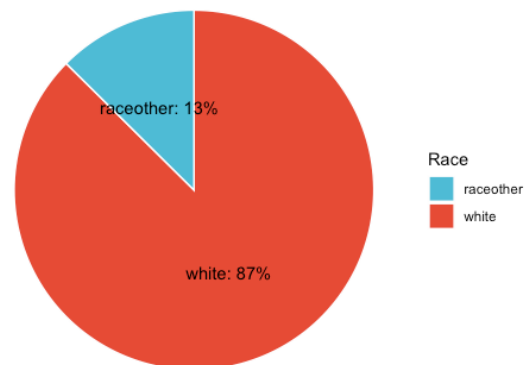

Figure 2: Racial distribution among deepfake targets on TikTok and YouTube

A possible explanation for the dominance of white targets would be that predominantly non-white nationalities use languages other than English to a higher degree than majority-white ones, and so would not be picked up by the crawler. However, few languages using the Latin alphabet have their own words for "deepfake" but use the English word, meaning that the crawler has likely picked up a somewhat even distribution of geographic origins. Still, white faces are greatly overrepresented as compared to internet users at large. Another explanation would be that the search is made from a US IP address, which could skew the hits to be more "relevant" to US users. And indeed, 45% of the targets are American. It is thus expected that the racial composition of deepfake targets resembles that of the United States. Even so, the white targets are hugely over-represented compared to the racial composition of the US, especially on YouTube. Although white people in general have far more access to institutions of power (both domestically in majority-white countries and globally), it does not seem that the deepfakes target white individuals in their capacity as white, but rather reflect an underlying asymmetry in terms of visibility.

In any case, the racial patterns are worth exploring further, especially with regards to how they compare to the pornographic sites. Although Home Security Hero (2023) provides some indications, such as the fact that 53% of pornographic deepfakes feature South Korean women, plenty of questions remain. Do white targets dominate in e.g., American pornographic deepfakes too? What is the explanation? And what does this say about the phenomenon at large?

3.4 Most common occupation

The third thing that stands out in the data is the professions of the targets, i.e., the industry or sector to which they belong. Perhaps unsurprisingly, the largest group are individuals in the entertainment industry, especially actors. On TikTok, 58% of targets are in the entertainment industry (of which the absolute majority are actors), 29% are politicians and the remaining 13% "other."[7] On YouTube, the share of entertainers is even more extreme, accounting for 75.5% of the content. This is largely due to the numerous videos featuring clips from movies where one actor's face has been superimposed onto another's body with captions like "What if Tobey Maguire starred in Marvel's spiderman movies?" (more on this below).

---

[7] The "other" category comprises a wide range of professions, including royalty, athletes and scientists, etcetera.

Proportions of professional sectors, TikTok deepfake targets
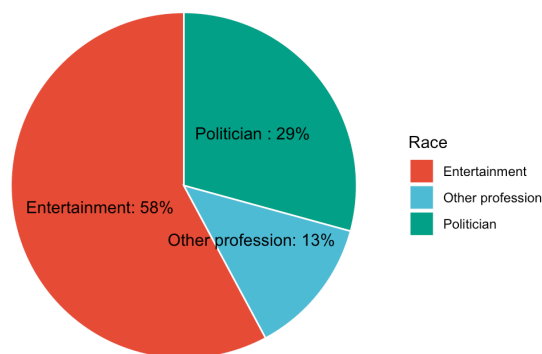
Proportions of professional sectors, YouTube deepfake targets
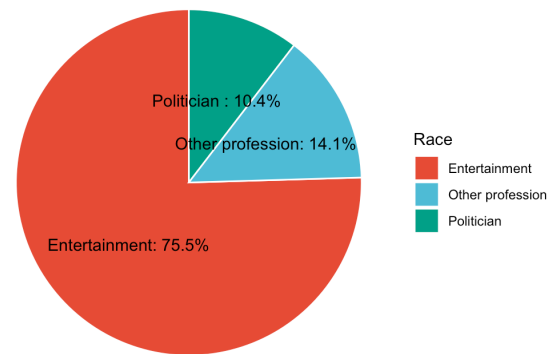
Figure 3: Occupational distribution among deepfake targets on TikTok and YouTube

The frequency of politicians, on the other hand, is harder to attribute to any specific type of video or trend. On YouTube, the presence of political deepfakes is so small that it hardly makes sense to draw any statistical inferences from it, but the distribution is 41% authoritarian, 47% liberal and 12% conservative. On TikTok, on the other hand, where political figures stand for a little less than one in three targets, 58% are classified as populist/authoritarian, compared to 39% liberal and only 4% conservative.

There are several possible explanations to the dominance of authoritarians. One is that individuals such as Donald Trump and Jair Bolsonaro are already associated with fake news and disinformation. Making a deepfake using their faces may thus be seen as an implicit or explicit comment on the absurd relationship to reality brought about by their admission to the realm of politics. A second, but related explanation, is that the fame of such individuals (especially Donald Trump) is based on, and intertwined with "internet culture" (e.g., 4Chan etc.) They are already commonly used in memes, and so their presence in deepfakes is merely a side product of them being part of an internet "lore." A third explanation is that male populist politicians exhibit a hyper-masculinity that can be mocked using deepfakes, for example by superimposing their face onto a female body. Female leaders like Marinne le Pen and Giorgia Meloni would have been possible targets, but appear only in one video on TikTok. Possibly, this is due to their lack of mockable masculinity. What strengthens this theory further is that the most frequent faces within the entertainer category, Arnold Schwarzenegger and Sylvester Stallone, arguably also exhibit a form of hyper-masculine persona. A more elaborate analysis of possible interpretations follows in the qualitative analysis.

3.5 Views

Though the frequency with which various target characteristics occur in the videos may inform our understanding of the supply of deepfakes, it is flawed as a means to measure their consumption. For example, about one in three TikTok deepfakes include politicians, yet these videos may, hypothetically, have a disproportional number of views, meaning that the consumption of the phenomenon is more political than indicated by the supply level alone.

A first look at the total number of views for each target/video characteristic indicates considerable variation (see Figure 4). For example, on TikTok, videos featuring male targets have a total of almost 560 million views, whereas women have only 144 million (not even one third of that of men). Videos featuring non-white targets on YouTube have a combined view count of 9.5 million compared to white targets, who total close to 30 million views—almost 2.4 times as many. When computing the average view count

for each characteristic, rather than the total, these differences are somewhat moderated, which suggests that the patterns are to some extent driven by a relatively small number of viral videos with tens of millions of views. Nevertheless, some notable variations remain. For example, videos featuring white targets on YouTube have an average of 1.6 million views, whereas videos featuring targets of other races have an average of only 0.48 million. Similarly, the difference between, e.g., videos featuring entertainers (1.2 million) and other professions (0.58 million) on TikTok is more than double.
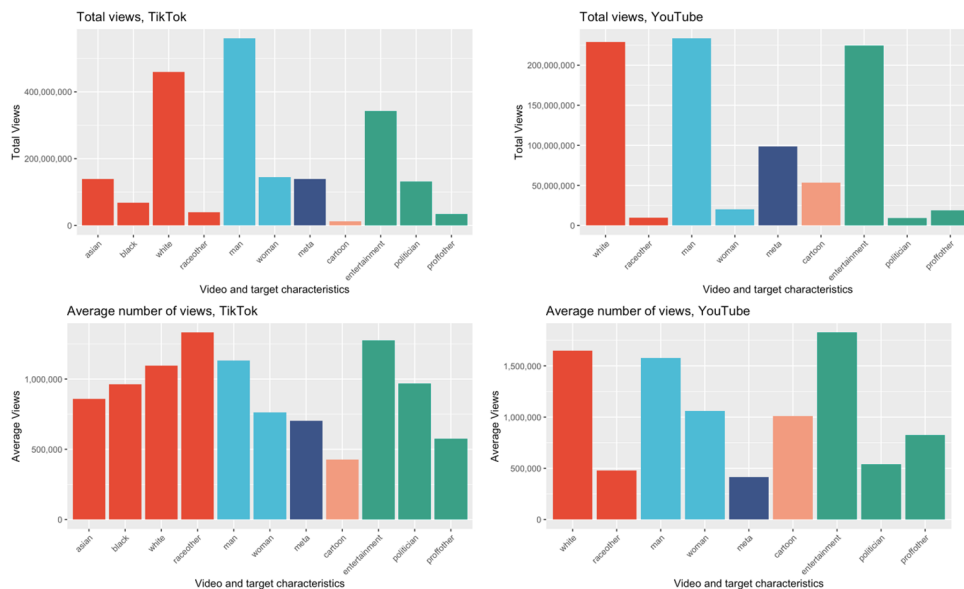

Figure 4: Variation in views across video and target characteristics

Overall, the patterns that emerge from looking at views is that white men within entertainment and politics dominate here too. The question is only to what extent the pattern is statistically significant, i.e., whether the target characteristics *drive* views. To test this, a multivariate regression analysis was performed using a T-test as a measure of correlation, and with views loged (see attached R code in the GitHub repository). Notably, almost none of the target characteristics had statistically significant effects. For TikTok, the only exception was gender, where videos starring male targets showed a positive T-value of 3.66 and a *p* value approaching zero. On YouTube, only *professional sector* came out statistically significant, with videos featuring entertainers displaying a T-value of 3.869 and *p* value approaching zero, and videos starring politicians displaying a T-value of 2.864 and a *p* value of 0.004.

The lack of statistically significant correlations should however not be interpreted as if the target characteristics do not matter for the extent to which a video is viewed. As indicated by the above plots, there is plenty of variation in the data. The result of the regression is probably rather due to the low number of observations within the minority characteristics. For example, there are only a handful of videos of women on YouTube. The big take-away is that videos featuring women and non-white targets are both rare *and* receive fewer views, although the former makes it difficult to compute the latter with statistical significance. In summary, taking views into account seems to confirm the patterns that emerged from merely counting the number of videos: the key targets of deepfakes on mainstream platforms are male, white and politicians or entertainers.

3.6 Interpretive analysis

In addition to the statistical analysis, a subset of 100 videos were also analyzed qualitatively so as to interpret their cultural meaning and context. Although more qualitative data, e.g., interviews or online ethnographic approaches, would be required to say something definitive about the motivations behind the

sharing and production of deepfakes, the present analysis provides an empirical contribution to the budding discussion on the cultural domestication of deepfakes as it (see e.g., Allison, 2021; Ayers, 2021; Glick, 2023; Holliday, 2021). Three cultural functions stand out stand out as particularly salient: (1) Augmentation of artistic imagination, (2) showing off technological savvy, and (3) mere amusement, or what may be interpreted as a form of *carnivalesque profanation.* Each is accounted for below, but extra focus is placed on the latter, as it arguably holds a key to understanding the production of deepfakes on deeper level.

3.6.1 Augmentation of artistic imagination
Strikingly, many videos in the sample actually appeared to aspire to a higher artistic value, e.g., engaging in discussions on the casting of Hollywood movies by "recasting" scenes by using the face(s) of the desired actor(s). As noted by Holliday (2021, 910), this type of video is commonly presented with explicit or implicit reference to Hollywood casting gossip. In the sample used in this study, this is especially common for franchises, like Spiderman, where the lead character has been played by multiple actors in the relatively recent past. But there are also other types of recasting videos, such as Indian channels that recast Hollywood movies with Indian actors (living and dead). Another example is a YouTube channel that appears to be exclusively dedicated to making deepfakes of Elvis Presley's face superimposed onto imitators. In the description to one of the videos, the channel's owner states:

> I would like to say that this is just a simple way to honor Elvis. It's impossible to reproduce the real Elvis in his completely unique personality. The goal is purely entertainment and reminding us of the great Elvis Presley. We can get closer using a more or less similar body model, but it will never be real, the real is long gone and only God could make another one like him. I think that what this technology already provides is something unthinkable even for our days.

As indicated by the quote (and indeed by the subsequent comments from viewers) the cultural function of these types of videos is not to mislead or show off, but to pay tribute to an idol. It is a means to augment one's imagination to help visualize what would have been possible were Elvis alive. Similarly, simulating what it would have been like, had this or that actor been casted in a given movie, are also a way of augmenting users' experiences of not only consuming but actively interacting with content in what Holliday (2021) calls a "take two" ramification. This kind of augmentation of creative imaginaries is reflected in a large number of videos in the dataset and may be a topic for further exploration in and of itself. It indicates that synthetic media is not just something that may disrupt the people consume entertainment, it is already happening.

3.6.2 Showing off technological savvy
In addition to the artistic function, the purpose of a great number of videos appears to be mere technological showing off in the sense of "look what is possible with this technology." There are entire channels dedicated to swapping faces of seemingly random actors in seemingly random scenes of various movies—all to reach maximally realistic results. On TikTok especially, some influencers even appear to exclusively post videos of themselves putting on the faces of various celebrities (at the request of followers). Since they do not engage in any particular activity once the face is superimposed, the main purpose is presumably just to show off one's technological skills and inspire awe at the wonders of technology. This is also aligned with the survey conducted by Home Security Hero (2023), which found that 57% of deepfake consumers attribute their interest in deepfakes to mere technological curiosity.

3.6.3 Carnivalesque profanation
The most common cultural function, however, is using deepfakes for comical effects. One of the most common tropes used to achieve such an effect is a middle-aged hypermasculine actor's face superimposed onto a female body (e.g., Arnold Schwarzenegger's face and voice on the body of Elaine from Seinfeld).

The contrast between the original and the superimposed face creates a comical effect which borders the absurd (see Holliday, 2021). Anyone could be anything—the borders are broken.

A great many videos featuring political targets fall into this category. Contrary to what is argued by Glick (2023), who sees a great satirical potential in deepfakes, most of the videos in the present data seem more focused on making political targets sing, dance or appear in funny TV shows. For example, one of the most prolific trends in the data is a reoccurring series of videos in which Joe Biden, Donald Trump and Barack Obama play Minecraft and talk over Discord alongside a series of more or less frequent high-profile guests like Joe Rogan, Ben Shapiro and Hilary Clinton. In these videos, the deepfake technology is applied only to the voices of the targets, which "reads" a script written by the video's creator, in which the presidents are portrayed as a group of friends who engage in harsh and profane banter. For illustration, consider this excerpt from one of the videos:

> Biden: Boys, are you down for a sleepover at the White House?
> Obama: I'm down bro.
> Trump: Me too.
> Biden: Ok, make sure to bring a tooth brush, a phone charger, some candy and a sleeping bag.
> Trump: Joe, it's the White House, why the fuck would we bring sleeping bags? We're not camping.
> Obama: Plus, don't you already have all that stuff provided?
> Biden: I was just kidding guys, that was a test.
> Obama: Do you guys wanna play Mario Cart at the sleepover?
> Trump: Yeah, I'm down bro.
> […]
> Trump: Okey, where should we order the pizza from?
> Obama: Of course, Trump the fat ass is already thinking about pizza.
> Trump: Shut the fuck up, we're getting Papa John's.

Here, as in the plethora of similar videos, there is no obvious political message beyond making fun of the individual characters. As noted by previous commentators, the use is *playful* and *comical* rather than provocative (Ayers, 2021; Holliday, 2021) To add a further qualification to the mere concept of "playfulness," I propose that the cultural function of the videos is rather to be interpreted as what Russian literary scholar Mikhail Bakhtin called *the carnivalesque* (1984), i.e., a challenge of hierarchies based on laughter, profanities and a general assault on stable identities.

The epitome of carnivalesque is the absurd comedy of Renaissance novelists like Rabelais, Cervantes and Bocaccio, which are filled with farts, excrements, sex and silliness. According to Bakhtin, this literary tradition is an extension of medieval festivals like the Feast of the ass and the Venetian Carnival, where hierarchies were temporarily put on hold to give room for the bodily experience of laughter and sex. Mock priests would wander the streets yelling profanities, men would dress as women, nobility as pleb, and so forth, all under the protection of masks that created a constant flux of identity. As elaborated by cultural historian James H. Johnson (2011, 184) in *Venice Incognito*:

> … masks encourage and affirm what the carnivalesque reveals. In place of dogma, certainty, and a fixed order on earth, the carnivalesque offers a universe that is open-ended and forever in flux. In the clarifying light of a world inverted, truth becomes relative. Masks convey to all that rank is arbitrary and status is only skin deep. Masks wean their wearers from conformity to the self and blur the line between reality and make-believe.[8]

The parallel to medieval carnivals may seem far-fetched, yet, when thus described, the role of the carnival mask bears a striking similarity to the way deepfakes are employed on today's video-sharing platforms. Much like carnivalesque masks, synthetic media is "blurring the line between reality and make-believe," not to actually persuade anyone of an alternate reality, but to provoke laughter and engage in collective

---

[8] For full clarity, Johnson is drawing on Bakhtin here.

profanity. One of its chief cultural functions is making the faces of quasi-sacred institutions like the US presidency and other semi-worshiped celebrities appear in contexts where they would never otherwise be seen.[9] Thus, the sharing of deepfakes on social media may be understood as a form of ongoing carnival, an heir of the Feast of fools, or the prose of Rabelais.

What does this say about the political significance of the phenomenon? To Bakhtin, the communal laughter and shapelessness of the carnivalesque held a revolutionary potential in revealing the arbitrary nature of hierarchy. As pointed out by Umberto Eco (1984), however, there is a distinction to be made between satire (humor with a political edge) and mere *clownishness*, where the carnivals of medieval and renaissance Europe arguably belong to the latter. Laughter is never politically potent when merely switching the roles of the powerful and the powerless, such as by profaning the faces of power through masks. Such shuffling of identities rather reinforces the universality of the oppressive structure. True satire, holds Eco, is always *dangerous* in that it challenges not only *who* holds power, but the structure which gives them legitimacy.

From the interpretive analysis of the videos in this study, it is clear that deepfakes on social media typically fall within the clownishness category. None of the analyzed videos appear to have a discernable political finesse. Whether this makes them harmless and politically impotent (as would be held by Eco), or whether this is what makes them an interesting cultural force (see Bakhtin) is not immediately clear. In fact, it is already a matter of debate. Ayers (2021, 1022), for example, has argued that the current swapping of identity merely reinforces masculine hegemony, noting that "creators take a seemingly apolitical, comical, 'safe' approach to identity rather than an explicitly critical stance." Holliday (2021), on the other hand, sees a far more radical potential and argues that deepfakes are "undoing" gender, and relieves the inherently performative nature of identity. Similarly, Glick (2023) points to the amazing possibilities of deepfakes imaginaries as a means to make "threatening and seemingly untouchable figures appear weak, vulnerable, and exposed" which can "help envision a more equitable and just future." While the present analysis pivots slightly in the direction of Ayers, a more thorough theoretical analysis is needed to say anything conclusive. The goal here is merely to provide empirical evidence for both sides to draw from.

3.7 Summary

Two research questions were put forth at the outset of this study. In response to the first one—*Who is featured in deepfakes shared on mainstream platforms?*—the above analysis shows that deepfakes typically feature white men from the entertainment industry or (less often) politics. Although a number of explanations for these patterns can be hypothesized, we ultimately need more qualitative analysis before saying anything with confidence as to why these patterns emerge and why some of them are so strong. To the second research question—*What is the cultural function of deepfakes shared on mainstream platforms?*—we may point to at least three key functions: (1) Augmentation of artistic imagination, (2) showing off technological savvy, and (3) *carnivalesque profanation.* It is important to recall that, while the analysis provided in this study suggests an interpretation of the content itself, it is unable to get to the actual motives of the creators. For this, further qualitative data, e.g., interviews with content creators, is needed.

4. Concluding comments

How do the findings relate to the larger literature on synthetic media? In one sense, they can be understood as a contribution to the growing pushback against the alarmist rhetoric surrounding the topic (see Habgood-Coote, 2023; Kalpokas & Kalpokiene, 2022; Kerner & Risse, 2021; Nieweglowska et al., 2023). As others have shown before, there is reason to treat narratives of an approaching "epistemic

---

[9] Although Donald Trump is perhaps, by his very persona, already a form of defilation of the sacredness of the presidential office.

apocalypse" with a healthy dose of caution, despite the seriousness of the proposed risks. The fact that there are currently plenty of deepfakes, with billions of views, on platforms that actively ban deceptive content, should add to this skepticism insofar as it points to alternate interpretations of the phenomenon's significance. In any case, it appears that the "high-profile" dramatic events (to use the words of Kalpokas & Kalpokiene, 2022, 4) that have dominated much of the literature (see e.g., Diakopoulos & Johnson, 2021; Schick, 2020b) are perhaps not the most urgent aspects to explore. It should be stressed, however, that the present study *does not attempt to say anything* about the proportion of deceptive vs. non-deceptive deepfakes on the studied platforms. Since the former is, by its very nature, impossible to quantify, the study can merely confirm a considerable presence of the latter.

A second contribution is about the demographics of deepfake targets. Previous studies of the faces featured in non-pornographic deepfakes, albeit not academic, have found a large presence of female targets. For example, the Home Security Heroes report (2023) finds that 77% of targets outside the pornographic sites are female. The report from Deeptrace (2019), which looks only at YouTube, finds 39% female targets. By contrast, the gender balance in this study is almost the reverse, with female targets in clear minority (28% on TikTok and 11% on YouTube). These discrepancies are possibly due to sampling strategies. Both of the previous reports namely identify deepfakes based on a list of channels dedicated solely to deepfake production, as opposed to going through search. While both approaches constitute legitimate sampling methods, the latter (as used in this study) almost certainly encompasses a broader variety. Notwithstanding the methodological differences, it can be stated that the strong dominance of female targets in the pornographic content is not reflected beyond the porn-specific sites. This strong male dominance warrants further inquiry into the cultural function of deepfake production.

Within the context of demographics, the strong dominance of white targets on both platforms, as well as the over-representation of authoritarians within the political targets also warrant further exploration. It is tempting to read these data as an indication that non-deceptive deepfakes represent a form of social critique or resistance since all the over-represented groups are associated with power. The interpretive analysis, however, does not support this. Although authoritarian politicians are common (especially due to the presence of Donald Trump), there is rarely any political finesse to the videos. They remain at the level of carnivalesque clownishness. Why authoritarians and whites are so over-represented is thus still an enigma in need of further investigation. There is clearly something going on here which cannot be explained by the present data alone.

Overall, the chief contribution of this study lies not merely in its corroboration nor questioning of previous literature. Rather, it lies in its broadening of the scope of the current debate by pointing to the rich, yet empirically unexplored, flora of non-deceptive deepfakes.[10] Given the seriousness of the possible disinformation threat, it is understandable that this (seemingly harmless) aspect has hitherto flown largely under the academic radar. Yet, if anything, the present study shows that the sheer size of the phenomenon, in combination with the enigmatic patterns of the target demographics, warrants further exploration. Indeed, now that a baseline knowledge about this side of the deepfake phenomenon is established, plenty of new questions arise. In particular, the study points to the need for more qualitative analysis of the communities and contexts surrounding the production and dissemination of deepfakes. What are the motives behind sharing them? If not sexual desire nor political sabotage, why do people spend hours producing them? How do they select their targets when not driven by sexuality? Answering such questions may not only give us a better understanding of the direction and potential of synthetic media generally, but may also inform the already heated debates regarding its deceptive qualities.

---

[10] While there is research on non-deceptive deepfakes (see Allison, 2021; Ayers, 2021; Glick, 2023; Holliday, 2021) it has largely taken a theoretical and (at most) interpretive angle, meaning that very little is empirically confirmed.

References

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). Report 2019: The State Of Deepfakes. September, 27.

Allison, T. (2021). Race and the digital face: Facial (mis)recognition in Gemini Man. Convergence, 27(4), 999–1017. https://doi.org/10.1177/13548565211031041

Ayers, D. (2021). The limits of transactional identity: Whiteness and embodiment in digital facial replacement. Convergence, 27(4), 1018–1037. https://doi.org/10.1177/13548565211027810

Bakhtin, M. (1984). Rabelais and His World. Indiana university press. https://doi.org/10.1215/00267929-30-4-605

Bhandari, A., & Bimo, S. (2022). Why's Everyone on TikTok Now? The Algorithmized Self and the Future of Self-Making on Social Media. Social Media and Society, 8(1). https://doi.org/10.1177/20563051221086241

Citron, D. K., & Chesney, R. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security National Security. HeinOnline. https://scholarship.law.bu.edu/faculty_scholarship/640

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. New Media and Society, 23(7), 2072–2098. https://doi.org/10.1177/1461444820925811

Eco, U. (1984). The frames of comic 'freedom.' In T. A. Sebeok (Ed.), Carnival! (pp. 1–10). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110848717.1

Glick, J. (2023). Deepfake Satire and the Possibilities of Synthetic Media. Afterimage, 50(3), 81–107. https://doi.org/10.1525/aft.2023.50.3.81

Godulla, A., Hoffmann, C. P., & Seibert, D. M. A. (2021). Dealing with deepfakes - An interdisciplinary examination of the state of research and implications for communication studies. Studies in Communication and Media, 10(1), 73–96. https://doi.org/10.5771/2192-4007-2021-1-72

Gosse, C., & Burkell, J. (2020). Politics and porn: how news media characterizes problems presented by deepfakes. Critical Studies in Media Communication, 37(5), 497–511. https://doi.org/10.1080/15295036.2020.1832697

Habgood-Coote, J. (2023). Deepfakes and the epistemic apocalypse. Synthese, 201(3). https://doi.org/10.1007/s11229-023-04097-3

Harris, D. (2019). Deepfakes: False Pornography Is Here and the Law Cannot Protect You. Duke Law & Technology Review, 17(1), 99–127.

Holliday, C. (2021). Rewriting the stars: Surface tensions and gender troubles in the online media production of digital deepfakes. Convergence, 27(4), 899–918. https://doi.org/10.1177/13548565211029412

Johnson, J. H. (2011). Venice Incognito : Masks in the Serene Republic. University of California Press. http://ebookcentral.proquest.com/lib/uu/detail.action?docID=669815

Kalpokas, I., & Kalpokiene, J. (2022). Deepfakes: A Realistic Assessment of Potentials, Risks, and Policy Regulation. Springer Nature. https://doi.org/10.1007/978-3-030-93802-4_7

Kerner, C., & Risse, M. (2021). Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds. Moral Philosophy and Politics, 8(1), 81–108. https://doi.org/10.1515/mopp-2020-0024

Maddocks, S. (2020). 'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes. Porn Studies, 7(4), 415–423. https://doi.org/10.1080/23268743.2020.1757499

Mashinini, N., & Africa, S. (2020). Criminal Liability for the Violation of Identity Using Deepfakes in South Africa. https://doi.org/10.34190/IWS.21.065

Nieweglowska, M., Stellato, C., & Sloman, S. A. (2023). Deepfakes: Vehicles for Radicalization, Not Persuasion. Current Directions in Psychological Science. https://doi.org/10.1177/09637214231161321

Öhman, C. (2019). Introducing the pervert's dilemma: a contribution to the critique of Deepfake Pornography. Ethics and Information Technology. https://doi.org/10.1007/s10676-019-09522-1

Perot, E., & Mostert, F. (2020). Fake it till you make it: An examination of the US and English approaches to persona protection as applied to deepfakes on social media. Journal of Intellectual Property Law and Practice, 15(1), 32–39. https://doi.org/10.1093/jiplp/jpz164

Schellewald, A. (2023). Understanding the popularity and affordances of TikTok through user experiences. Media, Culture and Society, 45(8), 1568–1582. https://doi.org/10.1177/01634437221144562

Schick, N. (2020a). Deepfakes : the coming infocalypse. In TA - TT - (First U.S.). Twelve. https://doi.org/LK - https://worldcat.org/title/1150998815

Schick, N. (2020b). Deepfakes : the coming infocalypse. In TA - TT - (First U.S.). Twelve. https://doi.org/LK - https://worldcat.org/title/1150998815

Sheffield, R. (2020, February). How Nipplegate Created YouTube. Rolling Stone Magazine.

Soukup, P. A. (2014). Looking at, through, and with Youtube. Communication Research Trends, 33(3), 3–34. https://scholarcommons.scu.edu/com

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11), 39–52. https://doi.org/10.22215/timreview/1282