



This is an AAM of: Wareham Morris, Katy (2024) Deepfake sockpuppets: the toxic 'realities' of a weaponized internet. In: Gothic Nostalgia: The Uses of Toxic Memory in 21st Century Popular Culture. Palgrave Gothic (PAGO) . Palgrave Macmillan, London, pp. 61-79. ISBN 978-3-031-43851-6; 978-3-031-43852-3 (eBook)

Available at: https://link.springer.com/chapter/10.1007/978-3-031-43852-3_5

Deepfake sockpuppets: the toxic 'realities' of a weaponized internet.

Katy Wareham Morris.

Deepfakes combine the latest in machine learning and Artificial Intelligence to create hyper-realistic audio-visual forgeries. Using state of the art technologies to engineer computer-generated, synthetic movement synchronised with fake audio tracks, deepfakes represent the next generation of digital media manipulation and disinformation: online propaganda cognitively more arresting than other text or image-based fake news, and which improve in quality with each new fake.

Industry experts and academics in the computer sciences have acknowledged the use of deepfakes to produce pornographic videos, fraud and fake news. This chapter investigates the pervasive, deceptive potential of deepfake 'sockpuppets', that is the creation of fake identities that generate and perpetuate toxic propaganda online. These 'sockpuppets' operate undetected as internet users do not have the forensic technology nor media literacy to identify the forgery. Even global news agencies and their editors do not have the expertise of global intelligence agencies who might be able to detect algorithmic glitches or weaknesses. This has meant that these non-existent, untraceable personae have manifested as international journalists generating publishable content, with the intention of distorting public perception and discourse. Perhaps more worrying, is that even on the discovery of the deepfake this 'news' is not retracted nor removed from websites.

This chapter investigates who is behind the 'sockpuppet' and their intentions. It argues that they are deployed as the cutting edge in toxic media propaganda to undermine cultural and political relationships and campaigns. This form of deepfake is purposefully provocative, working to compromise world leaders and remind nations of geopolitical conflicts. Through the use of xenophobic, racist and misogynist ideologies 'sockpuppets' incite social tensions whilst also undermining the internet as a source of reliable factual information. This has the startling potential to weaponize the internet, moving to a position where nothing is 'true' and maybe that doesn't matter.

Deepfake sockpuppets: the toxic “realities” of a weaponised internet.
Katy Wareham Morris.

Deepfakes combine the latest in machine learning and Artificial Intelligence (AI) to create hyper-realistic audio-visual forgeries. Deepfakes materialise from a specific type of deep learning – hence the name – in which sets of algorithms compete in a generative adversarial network or GAN (Goodfellow et al. 2014). Deep learning is a form of AI in which sets of algorithms (or neural networks) learn to simulate rules and replicate patterns by analysing large data sets. The generator algorithm creates content modelled on source material, which is already readily available (social media content, for example), whilst the discriminator algorithm seeks to detect flaws on the fake. This is an iterative process allowing for rapid improvements which address the flaws, meaning the GAN is able to produce highly realistic fake video content. This field of AI continues to advance at a rapid rate thus it is ever easier to create sophisticated and compelling videos. This has sparked concerns from politicians and international organisations about how these types of videos may be weaponised for malicious ends.

Early deepfakes manifested in the form of “false pornography” face-swap: female celebrities’ faces superimposed into sex videos without their consent (Harris 2019). Although social platforms reacted to this type of sexually exploitative material and banned it from platforms, the novelty of the technology has not gone away; face-swap images proliferate the internet, with this type of software available as free open-source downloads and hundreds of “how-to” videos on YouTube even a beginner is able to create and deploy a sophisticated deepfake.

Other forms of deepfake benefit from convincing audio track and lip synch and, real-time facial re-enactment often referred to “puppet-master” (Appel and Prietzel 2022). For this to work the data input to create the synthesised image involves head positions, eye movements and facial expression, for example typically generated by someone sat in front of a camera (Kim et al 2018). This type of deepfake is also paired with a “voice” either an impersonation or a fully synthesised imitation of the subject’s. When these audio-visual components are brought together through skilful post-production techniques the resulting video is incredibly convincing, almost undetectable. Agarwal et al (2019) call to attention that even with the aid of advanced forensic technology it may be almost impossible to identify this type of deepfake.

Deepfake technology has infiltrated various contexts: the media and creative industries, for example in the film industry (Golding 2021; Cameron 2021); in marketing and advertising (Kietzmann et al 2020); in galleries and museums (Mihailova 2021); as well as in political spheres and international relations (Chesney and Citron 2019). Coats (2019, p. 7) points out that in 2019 the U.S. intelligence community declared that opponents and competitors would attempt to use deepfakes to augment political campaigns, whilst Chesney and Citron (2019) acknowledge that deepfake applications have the potential to diminish democracies. This paints a scary picture which becomes even more noxious in the context of instantaneous digital dissemination through a variety of networks and platforms, afforded by web 3.0 functionality and user interaction. Much of the media content of the last century remained the prerogative of the trusted media

companies and professionally trained and qualified journalists. Yet social networking platforms have ensured that the definition of “news” and its circulation have changed (Turner 2010).

One of the symptoms of the postmodern condition is the public’s loss of faith in traditional ideologies or value systems. This feeling can be evidenced in the public’s attitude towards professionalism journalism also: audiences have lost faith and interest in the traditional forms of journalistic output (Waisbord 2018). This may be because there is increasing concern about the concentration of power in the mainstream media industries, whilst emerging platforms of journalism draw on the voices of the public itself. The asymmetrical distribution of power, that sees the media in control not only of the flow of information but effectively of the whole of the symbolic economy, is regarded as elitist, anti-democratic and as contributing to the running-down of the public sphere (Freedman 2018; Pickard 2018).

New and emerging platforms of journalism include various citizen and public journalism projects, which serve different political and commercial agendas. These may include the news blog, and “alternative” news sites which are just as happy to publish an interesting contribution from an amateur member of the public as a more conventional journalistic article from a freelancing professional. However, the increased volume of “news” does not necessarily mean that journalism is better, more reliable or more diverse in its political perspectives and sources of information. In fact, the opposite may be the case.

Vousoughi et al (2018) found that much of the content on social networking sites and messenger applications was unverified with people most likely to share negative or novel information; ominously false political stories were particularly effective in being spread. It is clear then that deepfakes contribute to the broader problems of fake news and dis-information which manifest within a post-truth society. Fake news blends traditional news elements with sensationalism, misinformation and bias (Mourão and Robertson 2019), whilst post-truth acknowledges the erosion of the very foundations of a liberal democracy, where the institutions which have previously constructed the world are disrupted due to declining trust in expertise combined with the rise in digital media (Harjuniemi 2021). Due to the unparalleled sophisticated realism, speed, scale and ability to personalise disinformation, deepfakes contribute to this “epistemic crisis” (Dahlgren 2018) with the potential to destabilise public engagement and therefore democracy.

The realities of a post-truth society also bring into question whether knowing and discovering a deepfake is even worth cogitating when everything can be perceived as biased and the traditional models of truth-telling are shattered (Waisbord 2018). The likelihood is that these videos will be watched and shared because they are deepfakes. This is because audiovisual content is more visually arresting Fallis (2021) and “seeing is believing” (Goldberg et al 2019). It is arguable that flaws in audiovisuals due to technological glitches such as asynchronicities, unnatural movements or rendering lag, may disarm the potential threat. However, the research shows that users ascribe accuracy to deepfakes even when imperfect (Kobis et al 2021; Dobber et al 2021).

As noted by Appel and Prietzel (2022), detection is also based on content and context, but when taken out of context as a result of peer-to-peer sharing and, when it is possible to personalise

disinformation (Diakopoulos 2019), the manipulation is even more insidious. Deepfake glitches may evoke suspicion but the nature of this technology demands that it becomes more sophisticated and less detectable with each new fake, whilst both content and context are likely to become more malleable and indeterminable on increasingly larger scales. Even when relevant information is available to detect the deepfake, individuals tend to rely on associative processes and intuitions, rather than engaging in arduous cognitive processing (Pennycook et al 2016) to make sense of what they're seeing. Perhaps this is also because of the current mainstream popularity of deepfake technology which is linked to free applications and entertainment (E.G. Face-swap montages videos on YouTube or the @deeptomcruise account on Tik Tok which has nearly 4 million followers): deepfakes are perceived as harmless "fun", something to marvel and playfully experiment with, connected to the participatory, creative nature of internet cultures. Whether political motivated or not the evidence is substantive: deepfakes have the potential to shape audience's thoughts and behaviour (Dobber 2021).

Deepfakes represent the next generation of digital media manipulation and disinformation: a form of online propaganda cognitively more arresting than other text or image-based fake news, and designed to improve in quality with each new fake. In order to begin to understand the full potential of this technology, we must, as suggested by Bode et al. (2021), investigate how deepfakes might align with and set in motion cultural divisions. One existing malicious strategy for online deception and manipulation likely to intersect with deepfakes is that of "sockpuppet" accounts.

Kats (2020) provides a succinct definition of sockpuppets: "a fictitious online identity created for the purposes of deception". He goes on to acknowledge that they are also sometimes referred to as "sybils"; this is the case particularly in academic literature (see the work of Yang et al 2014 and Morstatter et al 2021). This is not a new phenomenon, there is a long history of fake accounts on the internet. However, this particular type of deceptive account is recognised for their association with malicious intent and content and the number continue to grow. Sockpuppets are created in large numbers by individuals or groups and they have a particular reputation for creating false opinions.

These sockpuppet accounts engage in "online astroturfing" (Zerback et al 2021), a communicative strategy to create the impression that a particular opinion is widely supported. It is a deliberately persuasive form of disinformation, combining both media and political techniques. Manufactured user comments, under the guise of the sockpuppet account, are constructed to appear authentic and populate high traffic news sites and social networking platforms. Indeed, the audience are completely unaware that the sockpuppet user is biased, being sponsored, being malicious, or all of these (Zerback et al 2021). The astroturfing comments created and disseminated by sockpuppet accounts have several characteristics which make them appear realistic and believable: they are personal, which attracts audience attention and makes persuasion more likely (Taylor and Thompson 1982). The opinions articulated are planted within the mind and memories of the audience and are therefore more likely to affect judgment (Zillmann 1999; Domke et al 1998). These "people" appear to be fellow citizens, similar to other social media users, who are often perceived to be more trustworthy than politicians, for example

(Lefevere et al 2012); this perceived trust and similarity facilitate the persuasive effects (cited in Zerback et al 2021). The audience are more likely to accept what is being posted due to the realistic manifestation and lack of definitive framework of knowledge for making sense of individual posts, which is also applicable to digital content in general oscillating in the post-truth environment. Even if suspicions are raised, certain influences or agendas suspected, the audience cannot be entirely sure and have no definitive way of finding out.

This type of disinformation strategy has been acknowledged widely within the context of Western elections (Kovic et al 2018) with evidence of targeted campaigns directly connected to the 2016 presidential election in the United States, for example (Bessi and Ferrara 2016). Kats (2020) notes that the top-performing fake news stories at the time of the election generated more engagement than real news stories, with the fictitious stories amplified by sockpuppet accounts. It is therefore worth recognising how these sockpuppet accounts might work in tandem with deepfakes as a coordinated strategy to generate and perpetuate toxic propaganda online.

When thinking of the political threat of deepfakes, we might automatically think of how key figures or world leaders might be impersonated by “bad actors”. For example, in 2022 a deepfake of President Zelenskyy was released in which the subject can be seen and heard to tell troops to “lay down their arms”. Although it remains unclear who created and deployed the deepfake, it was boosted on Russian social media Facebook-equivalent, VKontakte and hackers were able to get it livestreamed on a Ukrainian TV channel, Ukraine 24. The deepfake created a convincing lip-synch, yet the video distorts President Zelenskyy’s head and overall proportions; despite this poor quality, the video still gained traction and world leaders, academics and legal professionals warned of the unknown harm that the future of this type of video might create.

When we think of how this type of video might work together with sockpuppet accounts this type of video becomes more subversive, even more threatening. Sockpuppet accounts have been used to praise, defend or support a person or organisation, as a way of manipulating public opinion. Research has shown that Russia’s elite, who are closely linked to the Internet Research Agency (IRA), are key sponsors of this type of activity (Zelenkauskaitė and Bualduccini 2017); the IRA being linked to sockpuppet accounts mimicking support for Russian policies (Kovic et al 2018). Sockpuppet accounts are also employed as a way of circumventing restrictions, suspensions or bans (often referred to be technologists as “block evasion”). In this example, sockpuppet accounts might be utilised by the TV station hackers as a way of maintaining attention on the livestream after the event; they may continue to share and draw attention to the deepfake even when it has been debunked; they may continue to question the news stories publicly debunking the video. These sockpuppet accounts make it more difficult to control the spread and engagement with such deepfakes, making it almost impossible to manage damage limitation.

Certain types of sockpuppet accounts are utilised as a coordinated strategy for strengthening certain views or opinions which can make the opposing viewpoint appear inane or corrupt. One sockpuppet account may be created purposefully to appear unintelligent even bigoted, whilst another sockpuppet account is utilised to refute the “obviously” ludicrous opinions of the instigator. In the case of the Zelenskyy deepfake, this type of coordinated strategy might work to

suggest that there is a strong and vitriolic anti-Russian sentiment, whilst the other accounts refute this overtly xenophobic attitude. The intended effect of this might work to undermine more reasonable and justifiable arguments made against the Russian invasion. In this sense, the sockpuppet accounts work to cast doubt and uncertainty, rather than proactively advocating for or against President Putin. Chavoshi et al (2016) found that sockpuppets post regularly with a sustained pattern of behaviour unlike regular users, the agenda is to create digital “noise”, streams of conflicting digital information and threads, which are difficult to navigate, trace and fact-check. This misinformation pollution, operating within the post-truth context defined earlier in this chapter, casts doubt on all content, with some users not concerned with finding the “truth” anyhow.

It is worth noting here how the deepfake works as an arresting audiovisual disinformation thread, whilst the sockpuppet accounts work to bolster the reach through masquerading as misinformation, as “debate” within the digital public sphere. This malevolent partnership has the potential to mask or deceptively account for technical glitches; further complicate content interpretation through increasing confusion and argument around the purpose, meaning and scope; and, further obfuscate the potential to debunk at the point of initial dissemination through deliberate social spread in order to purposefully decontextualise.

This type of toxic partnership contributes to the complex and problematic regulatory and legal frameworks which arguably do or do not apply to deepfakes, sockpuppet accounts and their combinations. Diakopoulos and Johnson (2021) point out that who to pursue and assignments of blame are often obscured because of a lack of transparency about who is responsible for what. Where did it come from? How do we know? Is the deepfake created and controlled by the same people or organisation as the sockpuppet accounts? How can this be evidenced? What is the main priority: the deepfake or the sockpuppet accounts? Which is more damaging? And the list of questions go on. It is alarming to consider such a powerful, anonymous and tactical enemy (or enemies). The potential of the deepfake sockpuppet combination amplifies the already identified intent of these types of malicious digital content; that is not to necessarily to convince the world of President Zelenskyy’s surrender but to propagate seeds of doubt, which may result in a longer term, more widespread feelings of fear and confusion.

One outcome, associated with deepfakes in particular, and likely to reach more sinister conclusions when affiliated with sockpuppet accounts, is that of the “liar’s dividend”. The non-partisan agency, Future Advocacy, who highlight the impact of deepfake technology, define this particular outcome as “a phenomenon in which genuine footage of controversial content can be dismissed by the subject as a deepfake, despite it being true”. This provides another “opportunity” for sockpuppet accounts to work in tandem with deepfakes: in this instance, the sockpuppet accounts could incite discussion and incredulity towards real digital content with the motive of casting doubt regarding its authenticity. In the example already used, this would take the form of astroturfing threads from sockpuppets intent on debunking the debunking of the Zelenskyy deepfake. Trump supporters have also been linked to sockpuppet astroturfing which aims to protect Trump by accusing the media of using deepfake technology to discredit or implicate Trump. In cases such as these, the sockpuppet accounts would utilise the form and

mode of address already established through previous posts to cast doubt without controversy or sensationalism, perceived simply as chat “typical” of this user. Both deepfakes and sockpuppets create uncertainty, skillfully manipulating information and data which is recognisable and often relatable to the average user. Sockpuppets enhance the potential harm of deepfakes by supporting the latent intentions through sustained and consistent astroturfing comments, which are made available across multiple digital channels and platforms. This exposure ensures that these messages are cognitively easy to retrieve thus increasing perceived validity (Tormala et al 2002), with the effect of blurring the boundaries between genuine, authentic content and that which is maliciously engineered.

Although this chapter has thus far discussed the potential of deepfakes and sockpuppets to compromise world leaders and undermine political campaigns, it is widely acknowledged by researchers in the field that the aim of these forms of digital disinformation is not easily identifiable; indeed there appears to be no central aim. The examples previously identified serve to remind audiences of geopolitical conflicts, but to also draw from memory historic cultural and social frictions in order to contextualise these narratives; to “help” the audience understand who might potentially be behind them, whilst concealing the true instigators. The consequence of these technologies then is to not prioritise one conflict over another, but to needle at existing, long-standing tensions and divisions. This content is officious, but also subtle; it is consistent and systematic without being “seen”; it is sophisticated yet flawed, which seems to increase its believability. To pervade the internet in this way, is to weaponise it.

We know that deepfakes will become more sophisticated and thus their partnership with sockpuppets is likely also to develop in complexity. Deepfake imagery of sockpuppets could literally bring these accounts to life. It is possible to create images of people who do not exist with control over emotions, hair, gender, age and skin colour (Karras et al 2018; Choi et al 2018). This deepfake sockpuppet does not need to be a recognisable person to be relevant or convincing. This chapter acknowledged the rise in citizen journalism, which confirms that deepfake sockpuppet analysts, commentators, journalists and supporters could all plausibly exist and attract followers. Rather than astroturfing commentary, the deepfake sockpuppets could start posting their opinions to Tik Tok or Instagram motivated to share their story, experience and opinions like anyone else. This potential makes it clear that weaponised content does not have to be overtly political, but can mimic grassroots activism. In this case deepfake sockpuppets do not have to harm or discredit real people by impersonating them, or by moaking controversial statements; the harm they cause is generated simply by being their “real” self.

There is now a well-known case of a deepfake journalist sockpuppet that was able to publish six international articles and editorials (Satter 2020). Even after an investigation by Reuters, the “person” behind the deepfake sockpuppet could not be identified. As one of the tactics highlighted earlier, the account assimilated to the context and audience appropriately: Oliver Taylor is Jewish university student who loves politics and coffee; he has a Gmail account and telephone number. These details next to the photograph of “Oliver” help to create the recognisable citizen which, as mentioned earlier, helps to construct a plausible representation; the deepfake sockpuppet Oliver Taylor, is just a regular guy, a political journalist interested in

anti-Semitism and Jewish geopolitical affairs. Reuters first became aware of Taylor when academic, Mazen Masri alerted them to an article written by Taylor published in an U.S. Jewish newspaper, in which Masri and his wife, Ryvka Barnard of being “known terrorist sympathisers”. The couple were shocked by this allegation and confused as to why they were being singled out by this man. Yet there is no record of Oliver Taylor anywhere, not even a digital footprint. The papers who published “his” work could not confirm his identity. Masri had suspicions about Taylor’s profile photograph which led to a sophisticated forensic investigation exposing it as a deepfake. Even when the papers removed Taylor’s work, they still received emails from the associated account asking for them to remain online on the understanding that many journalists use pseudonyms; need-less-to-say, Taylor could not verify his identity.

This example points to the potential damage of the deepfake sockpuppet marriage. Masri and his wife likely targeted because of their associations with Palestinian rights and liberation movements. The source of the “attack” still unknown. No egregious deepfake video of a known world leader making controversial claims. No obvious astroturfing threads from Taylor. Instead plausible, international journalism, almost imperceptibly pointing towards one of the most well-known, hotly debated, volatile geopolitical tensions. This demonstrates what’s to come in the future of disinformation propaganda: totally convincing, totally untraceable identities manifesting in various forms on various platforms. The case of Oliver Taylor is now over three years old; Oliver Taylor of today might be sharing his commentary as deepfake video. More advanced, large-scale coordinated astroturfing sockpuppet supporters might be working to corroborate Taylor’s story, even identity.

Whilst the Taylor case drew minimal social media attention, it still existed to distort public opinion and cast doubt on the integrity of young journalistic hopefuls looking to break into the industry. Today, it may have caused even greater harm to the targeted subjects, the audience and social institutions. Deepfake sockpuppets with the ease and freedom to operationalise various types of audio-visual and digital media forms will amplify these categories of harm. The technical capability to personalise deepfakes could lead to a profusion of “real”, unidentifiable sockpuppets activated through more narrow non-public (for example, chat) channels. This type of sharing has already been demonstrated through the harmless sharing of face-swap filter results, for example. In the future sockpuppet accounts will infiltrate chat groups and disseminate deepfakes either of “themselves” so as to create a conclusive “real” life or, of the unknown, entirely fabricated journalists or commentators they believe are worth sharing with their friends, as they somehow represent them.

As with memes (see Yoon 2016 and Williams 2020), this type of content is likely to replicate racist and misogynist stereotypes masked as parody or satire. As a site of ideological reproduction deepfake sockpuppets will surely contribute to the perpetual mockery of people of colour and women, denying structural racism and patriarchy through the forms and conventions of comedy. Not only that, but as with any media representation is it imperative to interrogate not only who is visible, but who remains invisible. There is a long history of racial bias in the film industry and recognised built-in bias in the field of AI (see Levy 2018 and Levin 2017). These value systems are often prioritised and embedded into the coding of these technologies as well as their outputs.

Deepfakes “read” human facial expressions, yet the people coding the algorithms are in the majority white and, the data training the algorithms are likely to be based on white faces that dominate media spaces. Ayers (2021) examined the practice of playful face swap videos and noted the dominance of white, male faces and noticeable lack of face swaps across racial categories. This reveals the persistence of historical pre-existing racial frameworks which reinforce boundaries for some identities over others, thus reinforcing the prevalence of white representation.

Moreover, deepfakes might be excused for exaggerating racial stereotypes by scapegoating a lack of technical efficacy. Whilst deepfake sockpuppets might perform stereotypical characteristics, emotions or personality traits which are then veiled by the “fact” that this is not an offensive caricature but a “real” person. Again, in both cases pre-existing social tensions are likely to be exacerbated. Bianchi et al points out that AI image generators are “extremely limited in the world they will create” (2022 p.10). Although it is possible to consider the potential deepfake technology might afford marginalised identities such as the LGBTQIA+ community, the discussions around valuable, democratised, personalised production might be optimistic, as the reality is deepfake production is labour intensive, time-consuming and demands skill; a deepfake sockpuppet even more so.

In 2023, popular podcaster Joe Rogan was the subject of a deepfake on TikTok; in the video Rogan boasts about a “libido booster for men”. He then gives information about the product and where to buy it. This deepfake is a significant example because Rogan has a following in the millions and represents a hegemonic masculine ideal. He is a white man with a muscular physique known for his passion for martial arts and UFC commentating; in the video he is facing the camera and he is clearly recognisable. Rogan is from New Jersey, a place stereotyped for being straight-talking; he is also of Italian heritage, the combination of the two being typically characterised as upholding strict patriarchal, heteronormative values. The topic of the video reminds the intended audience of stereotypical norms associating sex with a male prerogative, due to their naturally voracious appetite for it. Rogan is comedian well-known for his indelicate humour with sexual themes and innuendos; his reputation establishes that he is comfortable talking about sex. To further authenticate the video, Rogan’s colleague can also be seen in the video. It is clear that an endorsement from Rogan was intended to boost sales with a commercial intention, although men like this don’t typically confirm any sexual weakness in public. However, it is made legitimate through Rogan as role model, as spokesman for the type of men who identify with him, by drawing upon the stereotypical masculine ideals he epitomises: if Rogan needs to use this product and it works for him, it’s ok for me too and it’ll work for me. Earlier this chapter explained how perceived trust and similarity can make this type of communication more persuasive. Only it wasn’t Rogan, of course. The creators of this deepfake made a strategic choice when choosing to manipulate this footage.

The deepfake was posted and promoted by a TikTok user, whose account consisted of testimonial clips from supposed consumers. Not surprisingly this account is no longer accessible; most likely a sockpuppet account. Although the video was removed from TikTok and the original account banned, the deepfake was shared via Twitter and instantly went viral, this is despite the quite

obvious re-editing of existing podcast episode footage and fallible lip-synch. Rogan's colleague publicly debunked the video. This example exemplifies how deepfakes working strategically with sockpuppet accounts have the power to harness outdated cultural shorthands by means of creating convincing and persuasive content. Although not Political, this content is political and reinforces norms and expectations associated with gendered traits and behaviours. It also fortifies who is able to take up space in the public sphere: the opinions of white hegemonic men are valuable and valid. Rogan has been publicly criticised for his racist, misogynist views and for challenging known facts such as those associated with the COVID-19 vaccine, yet this reputation somehow chimes with the deepfake video content. The ideological signifiers are subtle, the representation is consistent with most of what is known and expected of Rogan, yet the possible questions around his own libido pique interest as the piece of the puzzle that doesn't quite fit. This engages the audience's attention: the sophisticated yet flawed manipulation of Rogan's popular reputation seems to increase its believability, even more so when it is recontextualised and rapidly disseminated via Twitter.

This is a good example of "persona plagiarism" (Diakopoulos and Johnson 2021, p. 2082), which focusses on the source rather than the content of a message. An unknown has manipulated Rogan's image and voice to create the TikTok deepfake, which serves the motivations and goals of the unknown. This video and therefore the motivations and goals too, are attributed to Rogan. However this is further complicated by the fact the deepfake was posted what was likely a sockpuppet account. As it impossible to know whether the TikTok account is genuine, credit it misdirected towards the well-known, recognisable celebrity, that being Rogan. Whether the video is received and interpreted positively or negatively is of no consequence as any outcome is misattributed to Rogan without his consent. Earlier this chapter drew attention to the fact that assigning blame is often undermined due a lack of accountability on the side of technology giants, geopolitical organisations and legal systems.

Earlier this chapter argued that deepfake sockpuppets manifest latent cultural and social frictions as a way of encouraging engagement. This example demonstrates how a popular and controversial figure like Rogan can be the victim of manipulated as a subtle political weapon, reminding the audience of what these types of men do and do not say, reinforcing well established toxic nostalgic narratives. Indeed, previous research confirms that current usage of deepfakes is currently overwhelmingly gendered and sexualised, with much of the harm directed toward women (Adjer et al 2019). Although a commercial incentive, there is arguably a subtle nod to the incel; consistent and systematic without being "seen". The content of the deepfake may be perceived as relevant to incel beliefs and interests and as such may be scrutinised more closely by that audience (Petty et al 1983). This type of deepfake sockpuppet represents Rogan as promoting a libido booster on TikTok for commercial reasons, which is further reinforced by the sockpuppet TikTok account pointing to who is behind the video yet acting as a decoy ensuring the real instigators remain unidentified. Perhaps most importantly, the audience are left questioning why a man like Rogan would need to use this type of product anyway, which subtly taps into the incel belief that the white male existence is under threat: if a man like Rogan needs this product then something must be wrong with the world.

This example also highlights the exploitative nature of deepfake technologies, where privacy is invaded for source material and the human image is commodified. Rogan's image, background and reputation are manipulated and perhaps defamed for commercial gain and consumer appeal without consent. It has been clear from the early days of deepfakes in pornography that more needs to be done to regulate and "police" creative integrity (Cook 2019): one should have the right to control their own face, defended as vehemently for the public as it for politicians, stars and celebrities. As the tools and applications of this technology become more sophisticated and even more widely available, creative, legal, ethical concerns around manipulation of performance and resemblance should be carefully deliberated. Whilst there also needs to be initiatives which improve awareness and understanding of the progressive and persistent nature of deepfakes due to rapid developments in the field of AI. Addressing media literacy will help with detection, enabling the public to identify technological glitches or at least pay closer attention. Specific deepfakes and associated sockpuppet accounts should be identified and made public so that people are able to see for themselves how the technology is able to improve and may go some way to "pre-bunking" future iterations. Although this increased knowledge might also serve to increase the potential of the "liar's dividend" whilst also contributing to the wider context of "truth decay" (Chesney and Citron 2019). As a result, all media content is to be doubted and social institutions are destabilised. We might also remind ourselves of the commercial gains to be had from these technologies, which might also provide a convincing rationale for why tough laws or policies and effective debunking strategies are prohibited.

This chapter has sought to investigate how deepfakes, sockpuppets and deepfake sockpuppets can harm the public and social institutions such as politics and the digital public sphere more generally, through fear, reputational misattribution and undermining trust. This chapter has explained how these technologies operate undetected as many internet users do not have the forensic technology nor media literacy to identify the forgery. Even global news agencies and their editors do not have the expertise of global intelligence agencies who might be able to detect algorithmic glitches or weaknesses. This has meant that these non-existent, untraceable personae have manifested as international journalists generating publishable content, with the intention of distorting public perception and discourse.

This chapter argues that deepfake sockpuppets are deployed as the cutting edge in toxic media propaganda to undermine cultural and political relationships and campaigns, whilst also undermining the internet as a source of reliable factual information. This has the startling potential to weaponise the internet, moving to a position where nothing is 'true' and maybe that doesn't matter.

List of References

Agarwal, Shruti, Farid, Hany, Gu, Yuming, He, Mingming, Nagano, Koki, and Li, Hao. 2019. "Protecting world leaders against deep fakes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 38-45.

Ajder, Henry Patrini, Giorgio, Cavalli Francesco, and Laurence Cullen. 2019. "The state of deepfakes: Landscape, Threats and Impact". Amsterdam: Deepttrace.

Appel, Markus. and Prietzel, Fabian. 2022. "The detection of political deepfakes". In *Journal of Computer-Mediated Communication*. 27:4, pp. 1-13.

Ayers, Drew. 2021. "The Limits of Transactional identity: Whiteness and embodiment in digital facial replacement" In *Convergence: The international Journal of Research into New Media Technologies*. 27:4, pp. 1018-1037.

Bessi, Alessandro and Ferrars Emilio. 2016. "Social bots distort the 2016 US Presidential election onlin discussion". In *First Monday*. 21:11.

Bianchi, Federico, Kalluri, Pratyusha, Durmus, Esin, Ladhak, Faisal, Chen, Myra, Nozza, Debora, Hashimoto, Tatsunori, Jurafsky, Dan, Zou, James and Caliskan, Aylin. 2022. "Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale". In *Computer Science*. 11 November, pp. 1-17.

Bode, Lisa, Lees, Dominic and Golding, Dan. 2021. "The Digital Face and Deepfakes on Screen". In *Convergence: The International Journal of Research into New Media Technologies*. 27:4. pp. 849-854.

Cameron, Allan, 2021. "Dimensions of the digital face: Flatness, contour and the grid. In *Convergence: The International Journal of Research into New Media Technologies*. 27:4. pp. 868-881.

Chavoshi, Nikan, Hamooni, Hossein, Mueen, Abdullah. 2016. "DeBot: Twitter Bot Detection via Warped Correlation". In *IEEE 16th International Conference on Data mining (ICDM)*. Barcelona: Spain. pp. 817-822.

Chesney, Robert., and Citron, Danielle, 2019. "Deep fakes: A looming challenge for privacy, democracy, and national security. In *California Law Review* 107, pp. 1753-1820.

Choi, Yunjey, Choi, Minje, Kim, Munyoung, Ha, Jung-Woo, Kim, Sunghun, Choo, Jaegul. 2018. "StarGAN: inified generative adversarial networks for multi-domain image-to-image translation". In *Proceedings of the computer vision and pattern recognition (CVPR)*. Salt Lake City: Utah. 18-23 June. <https://ieeexplore.ieee.org/document/8579014>. Accessed on 28th July 2023.

Coats, D. R. 2019. "Worldwide threat assessment of the US intelligence community". Senate Select Committee on Intelligence. Office of the Director of National Intelligence. <https://www.dni.gov/index.php/newsroom/congressional-testimonies/congressional-testimonies-2019/item/1947-statement-for-the-record-worldwide-threat-assessment-of-the-us-intelligence-community>. Accessed on 28th July 2023.

Cook, Jesselyn. 2019. "Here's what it's like to see yourself in a deepfake porn video". In Huffington Post. https://www.huffingtonpost.co.uk/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b0a3941861fced. Accessed on 28th July 2023.

Dahlgren, Peter. 2018. "Media, knowledge and trust: The deepening epistemic crisis of democracy". In Javnost – The Public. 25:1. pp. 20–27.

Diakopoulos, Nicholas. 2019. "Automating the News: How Algorithms Are Rewriting the Media". Cambridge, Massachusetts: Harvard University Press.

Diakopoulos, Nicholas and Johnson, Deborah. 2021. "Anticipating and addressing the ethical implications of deepfakes in the context of elections". In New Media & Society. 23:7, pp.2072-2098.

Dobber, Tom, Metoui, Nadia, Trilling, Damien, Helberger, Natali, and de Vreese, Claes 2021. "Do (microtargeted) deepfakes have real effects on political attitudes?" In The International Journal of Press/Politics. 26:1, pp. 69-91.

Domke, David, Shah, Dhavan.V. and Wackman, Daniel B. 1998. "Media priming effects: accessibility, association, and activation. In International Journal of Public Opinion Research. 10:1, pp. 51-74.

Fallis, Don. 2021. "The epistemic threat of deepfakes". In Philosophy & Technology. 34:4, pp. 623-643.

Freedman, Des. 2018. "Populism and media policy failure". In European Journal of Communication. 33: 6, pp. 604–618.

Future Advocacy. <https://futureadvocacy.com/> Accessed on 28th July 2023.

Goldberg, Matthew, van der Linden, Sander, Ballew, Matthew. T., Rosenthal, Seth. A., Gustafson, Abel, and Leisrowitz, Anthony. 2019. "The experience of consensus: Video as an effective medium to communicate scientific agreement on climate change. In Science Communication. 41:5. pp. 659-673.

Golding, Dan. 2021. "The memory of perfection: Digital faces and nostalgic franchise cinema". In Convergence: The International Journal of Research into New Media Technologies. 27:4. pp. 855-867.

Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron and Benigo, Yoshua. 2014. "Generative adversarial nets". In Advances in Neural Information Processing Systems 27, edited by Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, & K. Q. Weinberger. Cambridge, Massachusetts: MIT Press, pp. 2672-2680.

Harris, Douglas. 2019. "Deepfakes: False pornography is here and the law cannot protect you". In *Duke Law & Technology Review*. 17:1, pp. 99-128.

Harjuniemi, Hayek Timo. 2021. "Post-truth, fake news and the liberal "regime of truth": the double movement between Lippmann and Hayek. In *European Journal of Communication*. pp. 1–15.

Karras, Tero, Alia, Timo, Laine, Samuli, Lehtinen, Jaakko. 2018. "Progressive growing of GANs for improved quality, stability, and variation. In *Proceedings of the international conference on learning representations (ICLR)*. <https://arxiv.org/abs/1710.10196>. Accessed on 28th July 2023.

Kats, Daniel. 2020. "Identifying sockpuppet accounts on social media platforms". Norton Labs. <https://www.nortonlifelock.com/blogs/norton-labs/identifying-sockpuppet-accounts-social-media>. Accessed on 28th July 2023.

Kietzmann, J., Mills AJ., and Planagger K. 2020. "Deepfakes: perspectives on the future "reality" of advertising and branding. In *International Journal of Advertising*. 40:3, pp. 473-485.

Kim, Hyeongwoo, Garrido, Pablo, Tewari, Ayush, Xu, Weipeng, Thies, Justus, Nießner, Matthias, Pérez, Patrick, Richardt, Christian, Zollhöfer, Michael, and Theobolt, Christian. 2018. "Deep Video portraits". In *ACM Transactions on Graphics (TOG)*. 37, pp. 1-14.

Kobis, Nils, Dolezalova, Barbora and Soraperra, Ivan. 2021. "Fooled Twice – people cannot detect deepfakes but think they can". In *iScience*. 24:11. <https://doi.org/10.1016/j.isci.2021.103364>.

Kovic, Marko, Rauchfleisch, Adrian, Sele, Marc, Caspar, Christian. 2018. "Digital astroturfing in politics: definition, typology, and countermeasures. In *Studies in Communication Sciences*. 18:1, pp. 69-85.

Lefevere, Jonas, De Swert, Knut and Walgrave, Stefaan. 2021. "Effects of popular exemplars in television news". In *Communication Research*. 39:1, pp. 103-119.

Levin, Sam. 2017. "Google 'segregates' women into lower-paying jobs, stifling carers, lawsuit says". In *The Guardian* <https://www.theguardian.com/technology/2017/sep/14/google-women-promotions-lower-paying-jobs-lawsuit> Accessed on 28th July 2023.

Levy, Susan. 2018. "Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning". In *International Conference on Software Engineering*. May 2018, pp. 14-16.

Mihailova, Mihaela. 2021. "To Dally with Dalí: Deepfake (Inter)faces in the Art Museum". In *Convergence: The International Journal of Research into New Media Technologies*. 27:4, pp. 882-898.

Morstatter, Fred, Kim, Do Own (Donna), Jonckheere Natalie, Liu, Calvin, Seth, Malilka, and Williams, Dmitri. 2021. "I'll play on my other account: the network and behavioural differences of sybils" In *Proceedings of the ACM on Human-Computer Interaction*. 5: Chi Play, pp. 1- 18.

Mourão RR and Robertson CT (2019) Fake news as discursive integration: An analysis of sites that that publish false, misleading, hyperpartisan and sensational information. *Journalism Studies* 20(14): 2077–2095.

Pennycook, Gordon, Ross, Robert M., Koehler, Derek J. and Fugelsang, Jonathan. A. 2016. "Atheists and agostics are more reflective than religious believers: Four empirical studies and a meta-analysis". In *PloS One* 11. pp. 1-18.

Petty, Richard E., Cacioppo, John T., and Schumann, David. 1983. "Central and peripheral routes in advertising effectiveness: The moderating role of involvement. In *Journal of Consumer Research*. 10:2, pp. 135-146.

Pickard, Victor. 2018. "When commercialism trumps democracy: Media pathologies and the rise of misinformation society". In *Trump and the Media* edited by Boczkowski, Pablo J. and Papacharissi, Zizi. Cambridge, Massachusetts: MIT Press, pp. 195–201.

Satter, Raphael. 2020. "Deepfake used to attack activist couple shows new disinformation frontier". Reuters. <https://www.reuters.com/article/us-cyber-deepfake-activist/deepfake-used-to-attack-activist-couple-shows-new-disinformation-frontier-idUSKCN24G15E> Accessed on 28th July 2023.

Tormala, Zakary. L., Petty, Richard.E. and Briñol, Pablo. 2002. "Ease of retrieval effects in persuasion: a self-validation analysis. In *Personality and Social Psychology Bulletin*. 28:12, pp, 1700-1712.

Taylor, Shelley. E. and Thompson, Suzanne. C. 1982. "Stalking the elusive "vividness" effect". In *Psychological Review*. 89:2, pp. 155-181.

Turner, Graeme. 2010. "Redefining Journalism: citizen journalism, blogs and the rise of opinion". in *Ordinary People and the Media*. London: Sage Publications Ltd. pp. 71-97.

Vosoughi, Soroush, Roy, Deb, and Aral, Sinan. 2018. "The spread of true and false news online". In *Science*. 359:6380, pp. 1146-1151.

Waisbord Silvio. 2018, "Truth is what happens to news: On journalism, fake news and post-truth.". *Journalism Studies* 19(13): 1866–1878.

Williams, Apryl. 2020. "Black Memes Matter: #LivingWhileBlack With Becky and Karen". In *Social Media and Society*. 6:4, pp.

Yang, Zhi, Wilson, Christo, Wang, Xiao, Gao, Tingting, Zhao, Ben and Dai, Yafei. 2014. "Uncovering social network sybils in the wild". In *ACM Transactions on Knowledge Discovery from Data*. 8:1, pp.1-29.

Yoonn, InJeong. 2016. "Why is not just a joke? Analysis of internet mems associated with racism and hidden ideology of colorblindness". In *Journal of Cultural Research in Art Education*. 33, pp. 92-123.

Zelenkauskaite, Asta and Balduccini, Marcello. 2017. "Information warfare" and online news commenting: analyzing forces of social influence throughlocation-based commenting user typology" In *Social Media and Society*. 3:3, pp. 1-13.

Zerback, Thomas, Töpfl, Florian, Knöpfle, Maria. 2021. "The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them". In *New Media & Society*. 25:5. pp. 1080-1098.

Zillmann, Dolf. 1999. "Exemplification theory: judging the whole by the some of its parts". In *Media Psychology*. 1:1, pp.69-94.