TECHNOLOGY AS A THREAT OR A SOLUTION?
THE CHALLENGES OF RESPONDING TO SYNTHETIC MEDIA

Ignas Kalpokas

Associate Professor, Dr.
Vytautas Magnus University, Lithuania

Contact information
Address: V. Putvinskio g. 23, LT-44243, Kaunas, Lithuania
Phone: +370 37 327891
E-mail address: ignas.kalpokas@vdu.lt

Vaiva Šalaševičiūtė

PhD Candidate
Vytautas Magnus University, Lithuania

Contact information
Address: V. Putvinskio g. 23, LT-44243, Kaunas, Lithuania
Phone: +370 37 327891
E-mail address: vaiva.salaseviciute@vdu.lt

Monika Lipskė

Project Coordinator
Vytautas Magnus University, Lithuania

Contact information
Address: V. Putvinskio g. 23, LT-44243, Kaunas, Lithuania
Phone: +370 607 80443
E-mail address: monika.valentukonyte@vdu.lt

ABSTRACT

Synthetic media – defined as text, audio, images, and video content or entire 2D or 3D envi-
ronments generated by AI-enabled tools – are currently at the center of public attention. While
benevolent applications of such technologies abound, the negatives attract significantly more

debate. While some of such uses tap into existing fears of disinformation and related threats, others pertain to qualitatively new harms, such as non-consensual synthetic pornography. Of particular note is synthetic media's capacity to democratize content creation, for better or worse. Ultimately, such concerns lead to calls for policing synthetic media in terms of its automatic detection and removal. Nevertheless, such reliance on technological solutions has at least two undesirable effects: first, further concentration of power in the hands of online platforms and other technology companies and, second, ignorance of the underlying causes of nefarious uses of synthetic media. In this sense, generation of harmful content is best seen not as a standalone problem but as a symptom of underlying deeper – cultural – trends. As part of seeking a solution, this article traces some of the roots of nefarious synthetic content, ranging from non-consensual pornography to disinformation to toxic masculinity cultures and the insecurities atttendant to it.

KEYWORDS

Algorithmic governance, disinformation, synthetic pornography, technological solutionism, toxic masculinity.

INTRODUCTION

Synthetic media can be defined as digital content, primarily or exclusively produced using AI-enabled tools. Such content might include text, images, sounds, videos, or entire 2D and 3D environments. But humans are still involved in content generation by initiating the creative process (since AI, at least as of now, does not have an independent creative impulse), prompting the tools towards specific choices of subject and style etc., and evaluating the end result. The heavy lifting, however, is done by AI, thereby enabling individuals with limited skills and abilities to create artist-grade works or realistic depictions of what appears like actual people or events. While raising notable issues pertaining to, for example, Intellectual Property Law (as such generative models are usually trained on protected works alongside with content in the public domain)[1] and sparking debates as to the nature of art and creativity,[2] synthetic media nevertheless allow for new modes of self-expression and partaking in cultural production. Such synthetic media include generative models for images and text, deepfakes, and any other AI-enabled tools. In many ways, the pervasiveness of synthetic media is characteristic of today's mediatized societies, dominated by digital platforms and content and increasingly organized around their logics and affordances so that humans and digital tools become entangled in a mutually affective web.[3]

The increasing prevalence of synthetic media also raises significant security challenges. In fact, concerns about potential misuse of AI-generated content dominate already existing accounts in academia and beyond. In particular, the two generally positive aspects – the availability of tools and the ease of using them – are seen as dangerous and potentially democratizing disinformation and abuse, whereby even actors deprived of resources can pose serious threats. It is this threat discourse that also sparks the loudest calls to action and provision of solutions. Nevertheless, it is alleged that the focus on technological solutions to the problems posed by synthetic media is very likely misplaced as it is not only likely to fail in delivering the expected results but also misdirect attention from the underlying causes of the problems at hand.

The article is composed of three parts. The first part overviews the existing discourse on synthetic media, highlighting core concerns and challenges. The second part, meanwhile, provides a critique of the attempts to overcome the threats posed by synthetic media by way of technology, such as automated detection and removal tools; it is alleged that such technology focus succumbs to the solutionist fallacy by way of scraping the surface while providing a false sense of confidence nevertheless. Finally, the third part focuses on the potential to explain nefarious uses of synthetic media

---

[1] See e. g. Andres Guadamuz, "Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works," In *Artificial Intelligence & Intellectual Property*, edited by Jyh-An Lee, Reto Hilty and Kung-Chung Liu (Oxford and New York: Oxford University Press, 2021), 147-176; Martin Zeilinger, *Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property* (Lüneburg: Meson Press, 2021); Will Knight, "This Copyright Lawsuit Could Shape the Future of Generative AI," *Wired* (November 21, 2022a) // https://www.wired.com/story/this-copyright-lawsuit-could-shape-the-future-of-generative-ai/; Leigh McGowran, "'Legal Minefield': The Risk of Commercialising AI-Generated Images," *SiliconRepublic* (September 2022) // https://www.siliconrepublic.com/machines/ai-generated-images-legal-risks-copyright; James Vincent, "The Lawsuit that Could Rewrite the Rules of AI Copyright," *The Verge* (November 2022a) // https://www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data.

[2] See e. g. Arthur I. Miller, *The Artist in the Machine: The World of AI-Powered Creativity* (Cambridge (MA) and London: The MIT Press, 2019); Marcus Du Sautoy, *The Creativity Code: How AI Is Learning to Write, Paint and Think* (London: 4th Estate, 2020); Sofian Audry, *Art in the Age of Machine Learning* (Cambridge (MA) and London: The MIT Press, 2021); Oliver Brown, *Beyond the Creative Species: Making Machines that Make Art and Music* (Cambridge (MA) and London: The MIT Press, 2021); John Potts, *The Near-Death of the Author: Creativity in the Internet Age* (Toronto: The University of Toronto Press, 2022).

[3] Ignas Kalpokas, *Malleable, Digital, and Posthuman: A Permanently Beta Life* (Bingley: Emerald, 2021a).

through culture, particularly focusing on toxic masculinity. It thus transpires that while technological solutionism might provide an easy-to-embrace strategy for combating nefarious uses of synthetic media, such an approach does not solve the real underlying problem, which is the acceptability of generation and proliferation of harmful content within sections of society.

1. SYNTHETIC MEDIA AS A PROBLEM: THE THREATS DISCOURSE

It is a prevalent trend to perceive synthetic media as a threat which, in turn, implies a need for the society to be protected. The same applies to all of the major types of synthetic media discussed in this article: text, video, and images. Moreover, such media are not framed as mere simple threats but, instead, as ones shrouded in the clout of high-end technology and artificial intelligence. Another notable aspect is their techno-logically-induced ease of use, which in turn stokes fears of a deluge of manipulative content created by essentially anybody. All of these perceived attributes, as shown later, ultimately contribute to a search for equally technologically impressive responses with the hope of quick solutions becoming available.

Starting with AI text generators that produce text of a quality comparable to that written by humans, the main threat emphasizes in existing literature lies in the possibility that they can easily be used for nefarious purposes, including creating disinformation and powering social media bots in order to make the latter more effective by removing human operators.[4] It would be unwise not to admit to there being a threat – after all, experimental evidence suggests that people have a hard time distinguishing AI-generated text from that written by humans, so it can be used to simulate authentic opinions.[5] Moreover, with the introduction of more advanced and interactive tools, such as ChatGPT, the potential for creating nefarious text (both intentional and as a matter of bad training data or due to what's known as algorithm 'hallucination') increases significantly.[6] The same applies to another important type of synthetic content, namely, AI-generated images that have recently proliferated due to the emergence of powerful generators that necessitate as little as a text prompt. However, unless guardrails are built in, generation of disinformative and pornographic content is an unavoidable outcome.[7] While, at the time of writing, such generators still have quite a way to go towards achieving full and consistent photorealism of their outputs (although some, admittedly, come very close), reaching that milestone is only a matter of time and investment (although a powerful argument is to be made against photorealism – perhaps

---

[4]   Todd C. Helmus, "Artificial Intelligence, Deepfakes, and Disinformation: A Primer," *RAND Corporation* (July 2022) // https://www.rand.org/pubs/perspectives/PEA1043-1.html; see also Laura Illia, Elanor Colleoni and Stelios Zyglidopoulos, "Ethical Implications of Text Generation in the Age of Artificial Intelligence," *Business Ethics, the Environment & Responsibility* 32, no. 1 (2022): 4; Sarah Kreps, R. Miles McCain and Miles Brundage, "All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation," *Journal of Experimental Political Science* 9 (2022): 105.

[5]   Sarah Kreps, *supra* note 4, 106.

[6]   Dina Bass, "OpenAI Chatbot So Good It Can Fool Humans, Even When It's Wrong," *Bloomberg* (December 2022) // https://www.bloomberg.com/news/articles/2022-12-07/openai-chatbot-so-good-it-can-fool-humans-even-when-it-s-wrong?leadSource=uverify%20wall; Sharon Goldman, "The Hidden Danger of ChatGPT and Generative AI," *VentureBeat* (December 2022) // https://venturebeat.com/ai/the-hidden-danger-of-chatgpt-and-generative-ai-the-ai-beat/; Connie Lin, "How to Easily Trick OpenAI's Genius New ChatGPT," *Fact Company* (December 2022) // https://www.fastcompany.com/90819887/how-to-trick-openai-chat-gpt; Cade Metz, "The New Chatbots Could Change the World. Can You Trust Them?" *The New York Times* (December 2022) // https://www.nytimes.com/2022/12/10/technology/ai-chat-bot-chatgpt.html.

[7]   Rachel Metz, "Anyone Can Now Use Powerful AI Tools to Make Images. What Could Possibly Go Wrong?" *CNN* (October 2022) // https://edition.cnn.com/2022/09/30/tech/image-generating-ai-publicly-available/index.html.

lack of it would diminish the threat of manipulation and increase capacity for artistic self-expression). Still, even the current state of the art allows reasonably convincing manipulative content to be generated with ease, including simulating the appearance of actual individuals.[8] The latest development in AI generation – text-to-video generation – is typically treated in the same manner as the above: as a technological marvel but, simultaneously, a source of threat precisely due to its perceived technological sophistication.[9] With companies like Meta and Alphabet throwing their weight behind such tools, 'mainstreamization' of this type of synthetic video (alongside earlier forms of synthetic video, such as deepfakes) seems imminent.[10] Overall, discussions of such AI-based generative models tend to display an often uneasy balance between, on the one hand, freedom and the ability to express oneself and, on the other hand, potential for manipulative or outright harmful and illegal content.[11]

Still, most types of synthetic media, particularly text-to-image and text-to-video generators, are novel additions that have thus far attracted relatively little attention. Contrary to that, deepfakes have been around for a sufficient amount of time to accumulate a relatively large amount of critical attention. Also, more than any other type of synthetic media, deepfakes appear to be covered negatively almost by default, perhaps because there have been fewer well-publicized unequivocally positive creative examples since their introduction. Crucially, deepfakes allow users of the technology, for example, to superimpose the likeness of targeted individuals on images or videos of other people; when it comes to manipulated audio content, the same is done with the target individual's voice.[12,] Of course, synthetic media can have multiple humorous or self-gratificatory uses, such as inserting oneself into a clip from a Hollywood film or in order to parody somebody.[13] However, these same off-the-shelf apps can be used to maliciously portray others doing things they never did and provide 'proof' for conspiracy theories.[14] In fact, deepfakes are typically considered to be particularly potent vehicles of disinformation, resulting in polarization and segregation (particularly when strategically placed within susceptible echo chambers), erosion of societal trust, and incitement towards politically and ethnically motivated violence.[15] Moreover, the amount of audio and visual content that any social media user uploads to their profiles means that the threat of impersonation is not reserved to celebrities and public figures – in fact,

---

8   See e. g. Benj Edwards, "AI Image Generation Tech Can Now Create Life-Wrecking Deepfakes with Ease," *Ars Technica* (December 2022) // https://arstechnica.com/information-technology/2022/12/thanks-to-ai-its-probably-time-to-take-your-photos-off-the-internet/.

9   Mark Sullivan, "Meta's New AI Video Generator Could Be a Dangerous Misinformation Tool," *FastCompany* (September 2022) // https://www.fastcompany.com/90793071/metas-new-ai-video-generator-could-be-a-dangerous-misinformation-tool.

10   Spencer Feingold, "Artificial Intelligence Image Generators Bring Delight – and Concern," *World Economic Forum* (October 2022) // https://www.weforum.org/agenda/2022/10/ai-artist-systems-bring-delight-and-concern/.

11   Will Knight, "This Uncensored AI Art Tool Can Generate Fantasies – and Nightmares," *Wired* (September 2022b) // https://www.wired.com/story/the-joy-and-dread-of-ai-image-generators-without-limits/.

12   Todd C. Helmus, *supra* note 4, 3-4; see also Keir Giles, Kim Hartmann and Munira Mustaffa, *The Role of Deepfakes in Malign Influence Campaigns* (Riga: NATO StratCom COE, 2019).

13   Jan Kietzmann, et al., "Deepfakes: Trick or Treat," *Business Horizons* 63, no. 2 (2020): 7; Lucas Whittaker, et al., "All Around Me Are Synthetic Faces: The Mad World of AI-Generated Media," *IT Professional* 22, no. 5 (2020): 97.

14   Bart van der Sloot and Yvette Wagensveld, "Deepfakes: Regulatory Challenges for the Synthetic Society," *Computer Law & Security Review* 46 (2022): 1; see also Rob Cover, "Deepfake Culture: The Emergence of Audio-Video Deception as an Object of Social Anxiety and Regulation," *Continuum* 36, no. 4 (2022): 609; Molly Mullen, "New Reality: Deepfake Technology and the World Around Us," *Mitchell Hamline Law Review* 48, no. 1 (2022): 234.

15   Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 6; see also Laura Illia, Elanor Colleoni and Stelios Zyglidopoulos, *supra* note 4, 7.

most individuals are at risk.[16] Such risk is further reinforced considering the increasing number of easy-to-use apps that allow people with little to no experience to generate good-quality fake videos.[17] An even greater ease of use also applies to pre-trained text and image generators, since they follow commands written in natural language.

The democratic nature of synthetic content is in itself typically considered to be a threat, potentially resulting in the truth being drowned out '[i]n a world where any citizen has access to deepfake technology and can disseminate a fake video, photo or audio file on the Internet within seconds', offering content that is more novel and provocative than mainstream media would.[18] Similarly, as Robinson[19] asserts, it is not the technology as such that constitutes the threat but the ease of access to powerful tools: in fact, '[b]ecause the technology to manipulate media has always been there, it's just a matter of how easy and fast you can do this'.[20] Hence, the ease of use and increase of scale of manipulative output are seen as posing the biggest challenge. In this way, the effect of a disinformation campaign can be seen to significantly outweigh the likely costs, with the latter potentially being only marginal.[21] A simple cost-benefit analysis suggests that such a change of balance would make disinformation campaigns an ever more attractive weapon for an ever-growing array of actors, including resource-poor ones.

Particularly with the capacities of AI text generators in mind, it would be relatively easy to generate disinformative posts in bulk for a large number of fake profiles on multiple platforms simultaneously.[22] Meanwhile, with the advancement of text-to-video AI generators, malevolent actors could not only manipulate existing videos (which is typical of deepfakes) but also create completely new ones from scratch.[23] Such information threats are expected to be amplified through the creation of entire ecosystems of fake content generation, fake websites and social media accounts intended to spread it, fake news reports to give made-up content perceived credibility etc., making it extremely difficult 'to pierce through a multi-layered environment of deception'.[24] Synthetic media also permeate such ancillary tools: websites can be illustrated with AI-generated images and videos and filled with AI-generated text while fake social media accounts often benefit from deepfake profile pictures that cannot be traced as stolen or as stock photos, thus making such accounts significantly more difficult to identify.[25] In fact, there is even evidence to suggest that users are likely to find AI-generated faces to be more trustworthy than real human ones, perhaps because of how average and nondescript

---

[16]   Keir Giles, Kim Hartmann and Munira Mustaffa, *supra* note 12, 17; Lucas Whittaker, et al., *supra* note 13, 94–95.

[17]   Ruben Tolosana, et al., "An Introduction to Digital Face Manipulation," In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, edited by C. Rathgeb et al. (Cham: Springer, 2022), 3-26; see also Todd C. Helmus, *supra* note 4, 4.

[18]   Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 6; see also Maria Pawelec, "Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions," *Digital Society* 1 (2022): 3.

[19]   Rebekah Robinson, "AI Image Generators Enable the Creation of Fake Pictures to Support Fake News," *Coda* (October 2022) // https://www.codastory.com/disinformation/ai-image-generators-fake-news/.

[20]   See also Robert Chesney and Danielle Citron, "Deepfakes and the New Disinformation War: The Coming of Age of Post-Truth Geopolitics," *Foreign Affairs* 98, no. 1 (2019): 148; Don Fallis, "The Epistemic Threat of Deepfakes," *Philosophy and Technology* 34 (2020): 226.

[21]   Keir Giles, Kim Hartmann and Munira Mustaffa, *supra* note 12, 14.

[22]   Renee Diresta, "AI-Generated Text Is the Scariest Deepfake of All," *Wired* (July 2020) // https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all/; Bradley Honigberg, "The Existential Threat of AI-Enhanced Disinformation Operations," *Just Security* (July 2022) // https://www.justsecurity.org/82246/the-existential-threat-of-ai-enhanced-disinformation-operations/.

[23]   Mark Sullivan, *supra* note 9.

[24]   Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 2; see also Renee Diresta, *supra* note 22; Bradley Honigberg, *supra* note 22.

[25]   Todd C. Helmus, *supra* note 4.

they look.[26] Ultimately, it must also be stressed that the larger the amount of content (and profiles, real or fake) transpires to purvey a particular narrative, the more credible it is perceived to be.[27] In this way, ecosystems intended to disseminate and amplify synthetic (and potentially manipulative) content may well end up setting the agenda for public discourse courtesy of their sheer effectiveness in generating and propagating content.[28]

Even if such synthetic media-powered disinformation ecosystems fail to fully convince, they can simply jam the public sphere with noise, thereby making it significantly more difficult for citizens to understand what is happening on a specific issue and what the verifiable facts actually are.[29] In addition, danger is caused not only by the potential attempts at deception but also by the likely attempts to abuse the knowledge of such potential; the latter would likely lead to what is known as the liar's dividend, providing 'plausible deniability for any event caught on video'.[30] Hence, as Coeckelbergh[31] puts it, deepfakes and other AI-generated content 'may lead to a world in which it is no longer clear what is true and what is false, where facts and fiction mix'.[32] Under such circumstances, the most likely result would be not only an erosion of trust among citizens but also loss of perceived trustworthiness of political and knowledge authorities, be it the news media or scientists and other professionals.[33] Indeed, there is evidence to suggest that while deepfakes may not necessarily fully convince individuals, they can, at the very least, induce doubt and mistrust; beyond specific cases, events, and stories, this can lead to individuals becoming less likely to cooperate and engage in other forms of collective action, thus paralyzing civil society.[34]

In the environment described above, locking oneself within a partisan bubble would be an increasingly convenient solution for many citizens.[35] This would also mean that making political decisions becomes increasingly difficult while the legitimacy of any outcomes emanating from political processes would be severely compromised.[36] In addition, deepfakes are seen to pose a significant and acute threat to elections and democratic processes more broadly in terms of both growing polarization and new opportunities for malign domestic and foreign actors to manipulate information and frame political actors for misdeeds they had never committed.[37] In fact, the net result could then be a deterioration of societal order as such.[38] But there is also a further way in which society can be adversely affected, even without the need for high-profile political deception: instead, the real problem might lie in 'amplifying local harassment' of

---

[26] Sophie J. Nightingale and Hany Farid, "AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy," *Proceedings of the National Academy of Sciences* 119, no. 8 (2022): 2.

[27] Renee Diresta, *supra* note 22.

[28] Laura Illia, Elanor Colleoni and Stelios Zyglidopoulos, *supra* note 4, 5.

[29] Todd C. Helmus, *supra* note 4, 6; see also Bradley Honigberg, *supra* note 22.

[30] Lucas Whittaker, et al., *supra* note 13, 95.

[31] Mark Coeckelbergh, *AI Ethics* (Cambridge (MA) and London: The MIT Press, 2020), 103.

[32] For a similar argument, see also Ignas Kalpokas, "Problematising Reality: The Promises and Perils of Synthetic Media," *SN Social Sciences* 1, no. 1 (2021b): 8.

[33] Todd C. Helmus, *supra* note 4, 6–7; see also Samuel Woolley, *The Reality Game: How the Next Wave of Technology Will Break the Truth and What We Can Do About It* (London: Endeavour, 2020), 150; Bradley Honigberg, *supra* note 22.

[34] Christian Vaccari and Andrew Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," *Social Media + Society* 6, No. 1 (January-March 2020) // https://doi.org/10.1177/2056305120903408.

[35] William A. Galston, "Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics," *Brookings* (January 2020) // https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/.

[36] Maria Pawelec, *supra* note 18, 3.

[37] Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 6; see also Todd C. Helmus, *supra* note 4, 6.

[38] Samuel Woolley, *supra* note 33, 107.

activists and journalists by using synthetic media to adversely affect their reputations and credibility so that they do not challenge those in power.[39] Accountability – a necessary attribute of any people-oriented government – would be thus compromised.

Simultaneously, though, there might sometimes even be no need for some rogue actor to intentionally manipulate the public using synthetic media because this form of content, just as AI-produced results in any domain, is prone to algorithmic amplification of biases: after all, if training data contains biases (e.g. by way of learning the inequalities and discriminatory practices present in the society in question or if the tool has scraped through conspiracist or extremist websites and social media groups as part of its learning process), then such trends will be reproduced and even amplified in the output.[40] Indeed, the biases and learned toxicity in AI-generated content could be a widespread – but almost universally overlooked – problem with synthetic media. In this way, even content produced with no manipulative, disinformative, or otherwise harmful intent whatsoever (such as artistic and other cultural artefacts) can end up being severely implicated in the production of societally undesirable outcomes.

A further effect of note, and one that necessitates separate mention due to its pervasiveness, is the rise of so-called revenge porn, particularly against female teenagers and young adults as well as politicians, celebrities, and activists; the threat is significantly increased by synthetic media, and deepfakes in particular, as they make it possible for almost anybody 'to create a nude picture or porn video of someone based on a photo of her where she is fully dressed', leading to adverse societal consequences and distortion of the self-image of individuals targeted.[41] Indeed, the vast gender disparity in synthetic pornography (it featuring almost exclusively women) means that synthetic media 'exert a unique cost against women'.[42] Moreover, while most of the other nefarious use cases remain potential harms, the same cannot be said of synthetic pornography. The harm of the latter is both actual and widespread.

Overall, the harm landscape transpires to be rather significant. Again, it must be stressed that synthetic media are not devoid of beneficial uses; in fact, it could be claimed that synthetic media are rapidly becoming a new form of artistic expression. Nevertheless, the latter perspective is severely underrepresented in the dominant discourse on synthetic media. Moreover, this dominance of threat representations also implies a perceived need for a particular reaction – that of policing synthetic content. As the subsequent section shows, characteristically of today's platform-centric digital world, the response to a perceived threat is primarily imagined in terms of technological solutions that are expected to provide the society with the metaphorical silver bullet.

2. THE TECHNO-SAVIOUR: UNDERSTANDING THE SOLUTIONIST PIVOT

Currently, three ways of tackling the threats posed by synthetic media can be identified: a legislative approach, banning certain uses of synthetic media (such as non-consensual pornography), a technological approach, which focuses on identification and removal of synthetic content, and media and information literacy.[43] However, these approaches

---

[39]   Tom Simonite, "Forget Politics: For Now, Deepfakes are for Bullies," *Wired* (September 2019) // https://www.wired.com/story/forget-politics-deepfakes-bullies/.
[40]   Laura Illia, Elanor Colleoni and Stelios Zyglidopoulos, *supra* note 4, 3; see also Bradley Honigberg, *supra* note 22; Rachel Metz, *supra* note 7.
[41]   Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 7.
[42]   Todd C. Helmus, *supra* note 4, 7.
[43]   Anthony McCosker, "Making Sense of Deepfakes: Socializing AI and Building Data Literacy on GitHub and YouTube," *New Media & Society* (May 2022): 1–2 // https://doi.org/10.1177/14614448221093943.

cannot be separated: after all, any synthetic media-related legislation would be futile if there was no capacity to identify the offending content. Moreover, the development of generative tools means that synthetic content is increasingly difficult to detect without technological assistance.[44] This necessity for detection tools is also evident in case of private enforcement of platform rules, for example, in the attempts of stock image plat-forms, such as Getty, to ban AI-generated images: while the terms and conditions have been modified accordingly, in the absence of robust tools to detect and remove such images, the enforceability of the ban remains questionable.[45] And while some would label proposals for legislating on deepfakes or other regulatory attempts to the same or similar effect as 'alarmist',[46] the predominant point of view is that the likely or actual harms need to be mitigated at legislative or platform level or both.

Major social media platforms have also adopted their own policies with regards to detecting, monitoring, and policing synthetic content, particularly deepfakes, primarily focusing on content that has been specifically and intentionally manipulated so that it has the potential to mislead.[47] In order to combat potentially or outrightly harmful instances of synthetic media, and deepfakes in particular, AI-based detection systems have been in development almost since the advent of the technology itself, initially fo-cusing on detecting characteristic glitches, such as blinking abnormalities.[48] Likewise, major platform companies, such as Meta, are actively developing in-house and commis-sioning external tools for recognition and filtering of synthetic content.[49] Nevertheless, since synthetic content detection is often reminiscent of an arms race, ever new and more complicated methods have to be developed.[50] Significantly, it is the creators of synthetic media that have the first mover advantage because developers of detection tools can only attempt to catch up with what is already available.[51] The situation is even more complicated with regards to synthetic text, as it lacks most of the cues present in visual material; moreover, as generative models learn to mimic human writing style, and the conventional features of a particular topic and, for example, a given social me-dia platform, there are often few extraordinary features to go by.[52] Indeed, the fears pertaining to the release of ChatGPT, for example, include it being able to produce con-vincing text from academic essays to (often erroneous) medical advice to outright fake news articles in a way that makes it impossible to detect their AI-generated nature.[53] In

---

44   Maria Pawelec, *supra* note 18, 2–3.
45   Harrison Jacobs, "Getty Images Bans AI-Generated Images due to Copyright Worries," *ARTnews* (Sep-tember 2022) // https://www.artnews.com/art-news/news/getty-images-bans-ai-generated-images-due-to-copyright-1234640201/; James Vincent, "Getty Images Bans AI-Generated Content over Fears of Legal Challenges," *The Verge* (September 2022b) // https://www.theverge.com/2022/9/21/23364696/getty-images-ai-ban-generated-artwork-illustration-copyright.
46   Rob Cover, *supra* note 14, 616.
47   Molly Mullen, *supra* note 14, 233.
48   Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 2; Ruben Tolosana, et al., *supra* note 17.
49   Angel Vizoso, Martín Vaz-Álvarez, and Xosé López-García, "Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation," *Media and Communication* 9, no. 1 (2022): 295.
50   Todd C. Helmus, *supra* note 4, 10; Molly Mullen, *supra* note 14, 228.
51   Todd C. Helmus, *supra* note 4, 10.
52   Laura Illia, Elanor Colleoni and Stelios Zyglidopoulos, *supra* note 4, 4; see also Bradley Honigberg, *supra* note 22.
53   Ian Bogost, "ChatGPT Is Dumber Than You Think," *The Atlantic* (December 2022) // https://www.theatlan-tic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/; John Gapper, "ChatGPT Is Fluent, Clever and Dangerously Creative," *Financial Times* (December 2022) // https://www.ft.com/content/86e64b4c-a754-47d6-999c-fcc54f62fb5d; Rebecca Heilweil, "AI Is Finally Good at Stuff, and that's a Problem," *Vox* (December 2022) // https://www.vox.com/recode/2022/12/7/23498694/ai-artificial-intelligence-chat-gpt-openai; Alex Hern, "AI Bot ChatGPT Stuns Academics with Essay-Writing Skills and Usability," *The Guardian* (December 2022) // https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability; Connie Loizos, "Is ChatGPT a 'Virus that Has Been Released into the Wild'?" *TechCrunch* (December 2022) // https://tcrn.

fact, such texts may even raise suspicion by being too typical. Moreover, with synthetic text, there is no 'original' version to compare suspicious content against – in contrast to, for example, deepfakes of the face-swapping kind.[54]

Current detection tools are computationally expensive, meaning that there is a large financial (and environmental) cost in keeping them running all the time; however, if they are not running at full capacity in order to save on some of such costs, this creates loopholes for undesirable content to pass through the cracks.[55] The associated costs also mean that only the largest platforms can rely on having and deploying in-house tools.[56] Nevertheless, most of the available tools, including those based on machine learning, remain dependent on the specific training scenario that was used and, therefore, lack robustness when encountering previously unseen conditions'; their accuracy is further reduced in cases when content is compressed and resized when being shared on social media platforms.[57] Notably, despite the expenses, deepfake detection tools can currently only boost approximately a two-thirds success rate, and even when successful in identifying manufactured content, they are 'often are not able to indicate precisely what has been manipulated and how, let alone assess whether the manipulation was relevant or significant'.[58] Moreover, automated tools are incapable of understanding humor, particularly irony and sarcasm, and, therefore more than likely to misclassify such content as disinformative.[59] Another important aspect is the inability of technological tools to fully evaluate the context and to understand the intention of sharing specific content, such as understanding whether something is purposeful disinformation or mere misinformation, which is consequential with regards to freedom of expression.[60] Hence, both detection of manipulated content and, perhaps even more so, interpretation and evaluation of its intention and potential make technology-centric content governance much less straightforward than it would appear at first glance. Ultimately, involvement of human content moderators is still necessary,[61] which opens up the problems related to scale (the amount of content humans are capable of assessing) and bias/conviction, which AI-enabled tools are typically intended to eliminate. Nevertheless, the very presence of technological tools can give users a (relative) sense of security, which is a typical drawback of overreliance on the protective function of technology.

Another important challenge in deepfake detection is that that the scale of the task is asymmetrical. As Thomas[62] observes, "[w]hilst the good guys need huge numbers of deepfake videos to train on, the forgers might only need to place one video in the right place at the right time to achieve their goal". Moreover, deepfake detection is not only reactive in its nature but is also typically more time-consuming than the generation of fake content itself; as a result, deepfakes will probably remain undisproved for at least some time and, therefore, reach large numbers of people before their artificial nature

---

ch/3W7IZUY; Chris Stokel-Walker, "AI Bot ChatGPT Writes Smart Essays — Should Professors Worry?" *Nature* (December 2022) // https://www.nature.com/articles/d41586-022-04397-7.

[54] Renee Diresta, *supra* note 22.

[55] Juniper Lovato, et al., "Diverse Misinformation: Impacts of Human Biases on Detection of Deepfakes on Networks," *arXiv* (Accessed 13 December, 2022): 2 // https://arxiv.org/abs/2210.10026.

[56] Angel Vizoso, Martín Vaz-Álvarez, and Xosé López-García, *supra* note 49, 297.

[57] Ruben Tolosana, et al., *supra* note 17, 5.

[58] Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 2.

[59] Bradley Honigberg, *supra* note 22.

[60] Noémi Bontridder and Yves Poullet, "The Role of Artificial Intelligence in Disinformation," *Data & Policy* 3 (2022): 7.

[61] *Ibid.*, 8.

[62] Elise Thomas, "In the Battle Against Deepfakes, AI Is Pitted against AI," *Wired* (November 2019) // https://www.wired.co.uk/article/deepfakes-ai.

can be demonstrated convincingly.[63] Hence, even if detection and removal is successful, disinformative effects would still be experienced by those who had already encountered the content in question.[64] And, as long as different platforms will have divergent policies and technological capacities vis-à-vis synthetic content, the material in question will simply migrate laterally across the platform ecosystem to where it is least restricted.

It must also be stressed that while automated tools are incapable of distinguishing between the different use intention and cultural / societal value of synthetic content, a one-size-fits-all approach to policing synthetic media would be detrimental. After all, although positive uses of such technologies are often underrepresented in the literature, they are far from absent. Synthetic media can, for example, be used in consumer-facing services, offering maximum personalization[65] Additional uses may range from entertainment to arts and culture, enabling new forms of content and self-expression but also with clear commercial or educational applications, such as bringing dead actors or historical personalities back to life for e.g. films or teaching-oriented experiences.[66] Moreover, deepfakes may have a positive effect by helping remove language barriers, particularly in video content, by syncing lip movements in the source material with dubbed text and thus enabling broader cross-cultural distribution of commercial content or various awareness-raising initiatives.[67] No less importantly, identification of deepfakes is not always in everybody's best interest. For example, deepfakes can also be used to alter a person's appearance in order to safeguard the privacy of at risk or vulnerable individuals.[68]

Insofar as platform-based technological tools are concerned, there is always a danger of falling for what is often referred to as the temptation of technological solutionism – the idea that all problems can be solved by simply pouring in more data and applying new, more powerful, algorithms to crunch them or by any other piece of software.[69] In this way, complex societal problems are transformed into technological problems that can seemingly be addressed by merely tinkering with software or algorithmic tools.[70] Particularly in situations that require a rapid response (or are perceived as such), the idea of there being an app or other easy-to-use piece of technology becomes simply 'too attractive to ignore', regardless of their actual efficiency.[71] There is, notably, a rich background to solutionism: throughout Western intellectual history, technology has been perceived as a manifestation of human mastery over all kinds of 'others', from nature to much of humanity as well (those seen as 'primitive' or otherwise less capable). Hence, it should come as no surprise that once there is a need to regain a sense of perceived mastery, technology is turned to without much regard to

---

[63] William A. Galston, *supra* note 35.

[64] Angel Vizoso, Martín Vaz-Álvarez, and Xosé López-García, *supra* note 49, 296-297.

[65] Josh Constine, "ByteDance and TikTok Have Secretly Built a Deepfakes Maker," *TechCrunch* (January 2020) // https://techcrunch.com/2020/01/03/tiktok-deepfakes-face-swap/; Lucas Whittaker, et al., *supra* note 13, 97.

[66] Nicholas Diakopoulos and Deborah Johnson, "Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections," *New Media & Society* 23, no. 7 (2021): 2073; Don Fallis *supra* note 20, 226-227; Whittaker, et al., *supra* note 13, 96.

[67] Lucas Whittaker, et al., *supra* note 13, 96.

[68] Umur A. Çiftçi, Gokturk Yuksek and Ilke Demir, "My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymization," *arXiv* (Accessed 13 December, 2022) // https://arxiv.org/abs/2211.01361; Bart van der Sloot and Yvette Wagensveld, *supra* note 14, 3.

[69] Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (New York: Public Affairs, 2014).

[70] Luis F. Alvarez León and Jovanna Rosen, "Technology as Ideology in Urban Governance," *Annals of the American Association of Geographers* 110, no. 2 (2020).

[71] Linnet Taylor, "There Is an App for That: Technological Solutionism as COVID-19 Policy in the Global North," In *The New Common: How the COVID-19 Pandemic Is Transforming Society,* edited by Emile Aarts et al. (Cham: Springer, 2021), 209–215.

other considerations.[72] Nevertheless, it must be stressed that the technological solutionist approach operates by scratching the surface and optimizing the easily datafiable and quantifiable aspects of reality without regard to underlying conditions or biases potentially inherent in the datafication processes themselves.[73] Technological tools are also posited as ethically and politically neutral – which they, of course, never are in practice.[74] In this way, technological solutions enable the evasion of scrutiny that would otherwise be due.

There is a strong necessity for scrutiny, though, since technological solutionism also works by way of removing the boundaries between the public and the private, positing privately created (and often proprietary and inscrutable) tools as exclusive solutions to public matters.[75] Crucially, the automated nature of synthetic content detection and removal, just like any other technology-centric solutions developed by technology companies, pose challenges by operating in accordance with platform logics and business models, as opposed to the public interest; in this way, one can observe a shift from a free to an algorithmic marketplace of ideas.[76] It is thus increasingly difficult to separate public and private governance. Not only that: platform-centricity in governance ultimately ends up entangled within the problematic network of content policing and attention-focused content moderation. Hence, on the one hand, undesirable content needs to be filtered out but at the same time content that maximizes user attention and engagement is prioritized.[77] With synthetic media (although certainly not exclusively), those two imperatives can easily come into conflict.

The reliance on technological solutions to police synthetic content also ties in with the broader shift towards platform governance. As insightfully noted by Di Gregorio,[78] this shift has strong historical, ideological, and political roots: as the growth of digital technologies coincided with the neoliberal shrinkage of the state and market- and technology-optimism, the end result was, unsurprisingly, "accumulation of power by transnational private entities providing increasingly essential services". The result was also a transformation of the entire socio-technological architecture of the web – from an open-ended innovation-centric one to a much more platform-centric incarnation.[79] However, the influence of the rising prominence of platform governance goes beyond the digital world strictly conceived.

In fact, it can be observed that, due to their growing influence:

> online platforms also perform autonomous quasi-public functions without the need to rely on the oversight of a public authority; online public spaces are, therefore, effectively governed not by law (or, at least, not directly so) but by way of platform terms of service that offer technology companies both significant power and significant leeway to the extent that platforms 'enforce and balance individual

---

[72] Ignas Kalpokas, "Posthuman Urbanism: Datafication, Algorithmic Governance and Covid-19," In *The Routledge Handbook of Architecture, Urban Space and Politics Volume I: Violence, Spectacle and Data*, edited by Nikolina Bobic and Farzaneh Haghighi (London and New York: Routledge, 2023), 496-508.

[73] Siddharth Peter De Souza, "The Spread of Legal Tech Solutionism and the Need for Legal Design," *European Journal of Risk Regulation* 13 (2022).

[74] Luca Marelli, Katharina Kieslich and Susi Geiger, "COVID-19 and Technosolutionism: Responsibilization without Contextualization?" *Critical Public Health* 32, no. 1 (2022).

[75] *Ibid*.

[76] Giovanni Di Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (Cambridge and New York: Cambridge University Press, 2022).

[77] Ignas Kalpokas, *Algorithmic Governance: Law and Politics in a Post-Human Era* (London: Palgrave Macmillan, 2019).

[78] Giovanni Di Gregorio, *supra* note 76, 81.

[79] Niels Ten Oever, "This Is not how We Imagined It: Technological Affordances, Economic Drivers, and the Internet Architecture Imaginary," *New Media & Society* 23, no. 2 (2021).

fundamental rights by implementing automated decision-making processes outside any constitutional safeguard.[80]

Notably, platforms act not only as rule-givers – they acquire quasi-judicial functions as well; in fact, they "have showed that they perform functions which are similar to judicial powers and especially mirror the role of constitutional courts, namely the balancing of fundamental rights".[81] Moreover, in doing so, platform companies have very few obligations, particularly regarding human rights: after all, human rights protection documents have traditionally been targeted at states since they were, for most of recent history, the primary actors capable of threatening individuals.[82]

It is such platform-centricity that renders technological solutionism particularly dangerous. Firstly, focus on technology as a silver bullet only further increases the power of the largest technology companies that become tasked with policing large swathes of the public sphere, becoming rule-makers, rule-enforcers, and quasi-official proxies of states simultaneously. Secondly, the effects of something going wrong – algorithms 'misbehaving', such as making biased decisions or misclassifying content – are going to be felt even more strongly, particularly due to there being limited safeguards, including judicial ones. More likely than not, any other agents involved in the process would be reliant on the same digital tools, further limiting the possibilities of external arbitration.

A no less important problem, however, is that focusing on technological solutions does nothing to address the root causes of the nefarious uses of synthetic media. After all, despite the focus on the negatives, widespread in discussions of such content, there is nothing inherently sinister about synthetic media. However, should one care to look more deeply than the easily datafiable (and, therefore, technology-responsive), it transpires that many of the negative uses of synthetic media (such as generation of non-consensual pornography) manifest a strong cultural rooting and are, therefore, impossible to eliminate simply by way of technologically-enabled policing.

3. A CULTURAL TURN IN UNDERSTANDING SYNTHETIC MEDIA?

Techno-solutionist approaches to synthetic media largely ignore the cultural strata on which some of the nefarious uses are premised. Indeed, synthetic media are often presented as if they had appeared out of nowhere and unanticipated even though they, and deepfakes in particular, are "built on pre-existing practices and a longer history of digital creativity".[83] Instead of springing out of some kind of vacuum, innovation tends to follow emerging cultural practices, combining them with new technological capabilities; hence, it can be said that "deepfake technology has been developed over time in response to a perceived cultural need, desire or demand for the capability to generate footage that is as good as or better than recorded footage".[84] A similar cultural argument is also made by McCosker[85], for whom deepfakes represent a cultural adaptation to our increasingly datafied lives, implicating along the way 'platforms, code, and AI models in cultures of data production, cultivation, and use'. In this way, synthetic media can be seen as much more 'normal' than typical representations suggest. While cultural factors should not be taken as sole contributors to nefarious uses of synthetic media

---

[80] Giovanni Di Gregorio, *supra* note 76, 81–82.
[81] *Ibid.,* 118.
[82] *Ibid.,* 94.
[83] Rob Cover, *supra* note 14, 612.
[84] *Ibid.,* 612–613.
[85] Anthony McCosker, *supra* note 43, 2.

and, likewise, it is unlikely that one individual cultural factor could single-handedly explain the creation and distribution of aforementioned media, the very presence (and, it transpires, prominence) of such cultural strata demonstrates the limitations of technological solutionism in tackling undesirable synthetic content. In order to illustrate the role of cultural factors, this section focuses on the role of toxic masculinity and the associated technocultures.

On the one hand, this argument is open to criticism that the relative cultural 'normalization' of synthetic media, including its nefarious manifestations, simultaneously removes any critical sting of attempts to criticize the more egregious uses of the technology, such as non-consensual synthetic pornography, by extension rendering them just 'part of the culture'. Indeed, adherents to the cultural argument do seem to manifest a tendency of downplaying the problems and challenges posed by synthetic media. Notably, Cover[86] does so by way of ascribing to synthetic media the status of merely new cultural normality and dismissing any concerns raised about them as mere fear of uncertainty that always comes with something new. Such a dismissive attitude can certainly be seen as problematic and perhaps even insensitive. Hence, it is crucial to pay attention to the cultural antecedents of harmful uses of synthetic media without rendering them 'normal'. Nevertheless, it must also be stressed that merely inhering in culture does not automatically render something good and desirable or, at least, justifiable (and certainly not immutable).

Instead, of trivializing the harms caused by nefarious synthetic media, one could say that understanding the production and sharing of such content as a cultural phenomenon refocuses our understanding of emerging challenges and ways of dealing with them: from a technology-centric approach focused on finding and removing synthetic content (i.e. technological solutionism) to a more culture-focused one, posing a question as to what it is in the widespread assumptions and worldviews that, for example, makes individuals (mostly male) feel entitled to manufacture harmful synthetic content, particularly within the context of what is often referred to as 'toxic technocultures'[87]. In general, the latter can be defined as "the toxic cultures that are enabled by and propagated through sociotechnical networks".[88] The emergence of this cultural phenomenon is not new, reaching further both temporarily and contextually. As Massanari stresses, "toxic technocultures have always thrived in an environment of little accountability, anonymity, and the increased globalization enabled by online technologies".[89] Hence, not technology in isolation but its entanglement with culturally embedded toxicity becomes the real problem. Crucially, a cultural interpretation of synthetic media indicates that the disappearance of undesirable content is unlikely despite any attempts to regulate, ban, or detect and remove any such content since there are forces beyond mere technological affordances that motivate the existence and prevalence of nefarious uses of synthetic media.[90]

Firstly, generation of non-consensual synthetic pornography necessitates particular attention from the perspective of toxic cultures of masculinity.[91] In particular, the sense of entitlement and lack of empathy on behalf of those engaging in the generation

[86] Rob Cover, *supra* note 14, 615.
[87] Adrienne Massanari, "#Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures," *New Media & Society* 19, no. 3 (2015).
[88] *Ibid*., 333.
[89] *Ibid*., 334.
[90] Anthony McCosker, *supra* note 43, 4.
[91] See e. g. Olivia B. Newton and Mel Stanfill, "My NSFW Video Has Partial Occlusion: Deepfakes and the Technological Production of Nonconsensual Pornography," *Porn Studies* 7, no. 4 (2020).

of non-consensual synthetic pornography can be seen as stemming from "reliance on outmoded and poorly understood applications of evolutionary psychology, and a valorization of masculinity masquerading as a peculiar form of 'rationality'" that renders women and others not conforming to the 'male' stereotype as supposedly of lesser value.[92] For this reason, the perpetrators typically fail to adequately consider (or even admit the importance of considering) the effects of their actions on individuals other than themselves and members of their own (toxic) community.[93] In fact, it becomes evident that technological capacity and prowess, such as the ability to generate non-consensual synthetic pornography, almost by definition becomes "a way to access hegemonic masculinity with the sense of a right to do anything with technology, regardless of who might be impacted", thereby creating a toxic cultural mix in which harmful deepfakes and other kinds of nefarious synthetic media thrive'.[94] The psychological and symbolic violence thus exerted also provide 'an opportunity to perform the hypermasculine (through aggression and violence) without possessing any of the associated physical qualities'.[95] It is also such a dominance-oriented perspective that creates the sense of entitlement to order-keeping and boundary policing by the adherents of such views by way of generating and posting non-consensual pornography, thus demarcating (online) spheres and communities as their 'own'.[96]

Nevertheless, even with regards to disinformation and conspiracy theories, a notable connection can be traced: the use of such emotionally charged counter-mainstream content might, in fact, function as a rod to release the frustrations and insecurities of toxic masculinity.[97] More broadly, actual or perceived threats to hegemonic masculinity tend to invoke a strong and aggressive backlash, including by way of degrading others who do not conform to the view of male dominance or those seen as opposed to the institutions that have traditionally upheld such dominance.[98] That is particularly the case since the typical traits of toxic masculinity include a particularly pronounced domination drive as well as the ensuing competitiveness and insensitivity to others; those manifesting toxic masculine traits would, correspondingly, do anything to avoid defeat and the feeling of subservience, including resorting to disinformation and boundary-keeping through other forms of harmful content.[99] Especially in the context of "anger toward historical-cultural shifts in intersectional roles and identities across race-ethnicity-class-gender-sexuality-religion" as felt by some segments of the population, resorting to the rejection of such change by way of resorting to both disinformation (as an escape) and denigration of others (as projection of imaginary power) can be seen to gain traction.[100] Crucially, the proliferation of toxic content in online communities also helps to maintain trans-platform solidarity (thus signaling the presence

---

[92]   Adrienne Massanari*, supra* note 86, 333.

[93]   Olivia B. Newton and Mel Stanfill, *supra* note 91.

[94]   *Ibid*., 400.

[95]   Anastasia Salter and Bridget Blodgett, *Toxic Geek Masculinity in Media: Sexism, Trolling, and Identity Policing* (New York: Palgrave Macmillan, 2017), 8.

[96]   See Adrienne Massanari, *supra* note 86; Olivia B. Newton and Mel Stanfill, *supra* note 91; see also Daniel G. Heslep and P. S. Berge, "Mapping Discord's Darkside: Distributed Hate Networks on Disboard," *New Media & Society*, online first // DOI: 10.1177/14614448211062548.

[97]   See, notably Jayson Harsin, "Toxic White Masculinity, Post-Truth Politics and the COVID-19 Infodemic," *European Journal of Cultural Studies* 23, no. 6 (2020).

[98]   Debbie Ging, "Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere," *Men and Masculinities* 22, no. 4 (2019); Verity Anne Trott, "Gillette: The Best a Beta Can Get. Networking Hegemonic Masculinity in the Digital Sphere," *New Media & Society* 24, no. 6 (2022).

[99]   Jayson Harsin, "Aggro-Truth: (Dis-)trust, Toxic Masculinity, and the Cultural Logic of Post-Truth Politics," *The Communication Review* 24, no. 2 (2021).

[100] *Ibid*., 158; see also Adrienne Massanari, *supra* note 86.

of a broader toxic technoculture rather than community-specific views) as well as identity preservation and recruitment.[101] Overall, then, disinformative synthetic content, as understood from a perspective of toxic masculinity and the associated technocultures, entails community-building, escapism (building a synthetic world in which changes, perceived as threatening, have not taken place), and a fantasy of revenge (supposedly giving those perceived as an enemy what they deserve).

In this way, both non-consensual synthetic pornography and at least some manifestations of disinformation would represent a sort of unholy alliance between, on the one hand, postmodern creativity and mashup-oriented online cultures and, on the other hand, toxic masculine technocultures and a broader sense of "aggressive white male nostalgia for 'better' more privileged days".[102] Hence, synthetic media might provide a way for members of toxic technocultures to project imaginary power and dominance on a world in which they implicitly feel insecure and out-of-place.[103] Crucially, then, no amount of technological solutionism would eliminate this kind of content since any tools would be blind to the underlying root causes of harmful behaviors. Instead, efforts to tackle nefarious uses of social media should better focus on the attempts to tackle the cultural strata and, in particular, the toxic communities in which such harmful content thrives and even is mutually encouraged. However, this strategy cannot be interpreted as yet another simple and quick solution (as it is with the supposed technological fixes). It must be kept in mind that such communities are highly committed and have a tendency to respawn immediately.[104]

Of course, the very term 'toxic masculinity' is itself not without its critics as, according to some, it allegedly frames men as victims of the culture they are in rather than positioning them as agents who choose to be toxic themselves.[105] Others, meanwhile, criticize the idea of toxic masculinity from a very different perspective: for them, to the contrary, the term is too restrictive and allows men in general shed away responsibility.[106] However, focus on the individual level (the first criticism) would leave any analysis blind to the networks (including online) that reaffirm toxic identities, help collectively create and disseminate ideas and expressive tropes, and partake in further radicalization.[107] Alternatively, moving towards catch-all ascriptions (the second criticism) is unhelpful either, because it distracts from specific concentrations of undesirable behavior. In the end, it transpires that the reasons behind highly undesirable societal behaviors are complex and pertain to societal conditions that are difficult to datafy and, therefore, address by way of technological solutionism. The best that can be expected from the application of digital tools is policing of the effects and expressions of underlying societal problems. However, as long as the supply of such content is left unaddressed, a steady stream of harm will continue flowing while the communities involved in its generation would likely become even more radicalized.

It would certainly be overly ambitious to claim that cultural factors offer a singular explanation of the spread of nefarious synthetic media. Likewise, it would be patently

---

[101] Daniel G. Heslep and P. S. Berge, *supra* note 97.

[102] Jayson Harsin, *supra* note 100.

[103] See e. g. Emily van der Nagel, "Verifying Images: Deepfakes, Control, and Consent," *Porn Studies* 7, no. 4 (2020).

[104] See e. g. Daniel G. Heslep and P. S. Berge, *supra* note 97.

[105] See e. g. Andrea Waling, "Problematising 'Toxic' and 'Healthy' Masculinity for Addressing Gender Inequalities," *Australian Feminist Studies* 34, no. 101 (2019).

[106] Carol Harrington, "What is 'Toxic Masculinity' and Why Does it Matter?" *Men and Masculinities* 24, no. 2 (2021).

[107] Debbie Ging, *supra* note 89; Jayson Harsin, *supra* note 88; Verity Anne Trott, *supra* note 89.

wrong to assume that a single cultural strand, such as toxic masculinity (and associated technocultures), is *uniquely* responsible for actual and potential wrongs. Cultural factors may indeed contribute significantly to the rise of synthetic pornography and constitute one of the factors behind fake news and disinformation, but certainly not an exclusive one. What this brief discussion is aimed at, though, is an illustration of the complexity of the underlying strata and the multifaceted nature of individual motivations behind nefarious use of synthetic media. While it is beyond the scope of a single article to give a more comprehensive view of the cultural antecedents of such problematic behavior, the example presented here nevertheless represents a first step towards problematizing the alleged universality and straightforwardness of technological solutionism as applied to combating the harms of synthetic media.

CONCLUSIONS

The use of synthetic media is not always nefarious and, in fact, might well lead to an unprecedented growth in creative output and new forms of self-expression – although not necessarily art; just like, despite photography being an established art form, not everyone capable of taking decent photos with their smartphone is considered an artist. However, the type of use that attracts more attention than anything else is spreading disinformation and directly harming certain groups, such as targeting women with non-consensual synthetic pornography. While the capacity to create new types of harmful content is justifiably seen as a threat in itself, it is the ease and efficiency of such generation, whereby any individuals, including those lacking specific skills, can create content in bulk with the use of AI-enabled generative tools, often by way of little more than typing a prompt in natural language, that is seen as the ultimate gamechanger.

As is typical, the threat discourse also implies a search for solutions. While some of the proposed solutions are regulatory, ranging from legislation to platform-initiated bans, it is the technological aspect that ends up being emphasized. To an extent, this is only logical: after all, even regulatory solutions would be futile without the capacity to detect and remove the offending content. Nevertheless, not only such tools often lack effectiveness but they are also prone to the solutionist fallacy: the thinking that problems can be solved simply by collecting more data and applying more powerful analytics. Here the solutionist thinking has been demonstrated to have at least two problematic aspects: first, it only further increases the role and power of platform companies in regulating, policing, and adjudicating the public sphere and, second, it obscures the root causes of the problem by simply concentrating on the effects. Instead, it has been demonstrated that nefarious applications of synthetic media have notable cultural antecedents, including toxic masculinity. While the latter does not provide an exhaustive explanation of the root causes, it nevertheless demonstrates that one needs to look beyond algorithmic policing for meaningful solutions to nefarious uses of synthetic media. Hence, despite the appeal and the relative ease of technological solutionism, the oversimplifications involved are too significant for such policing of the digital public sphere to be viable, and the significantly murkier domain of cultural factors must be taken into account as well.

BIBLIOGRAPHY

1. Alvarez, León, Luis F. and Rosen Jovanna. "Technology as Ideology in Urban Governance." *Annals of the American Association of Geographers* 110, no. 2 (2020): 497–506.

2. Audry, Sofian. *Art in the Age of Machine Learning*. Cambridge (MA) and London: The MIT Press, 2021.

3. Bass, Dina. "OpenAI Chatbot So Good It Can Fool Humans, Even When It's Wrong." *Bloomberg* (December 2022) // https://www.bloomberg.com/news/articles/2022-12-07/openai-chatbot-so-good-it-can-fool-humans-even-when-it-s-wrong?leadSource=uverify%20wall.

4. Bogost, Ian. "ChatGPT Is Dumber Than You Think." *The Atlantic* (December 2022) // https://www.theatlantic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/.

5. Bontridder, Noémi and Yves Poullet. "The Role of Artificial Intelligence in Disinformation." *Data & Policy* 3 (2022): 1–21.

6. Brown, Oliver. *Beyond the Creative Species: Making Machines that Make Art and Music*. Cambridge (MA) and London: The MIT Press, 2021.

7. Chesney, Robert and Danielle Citron. "Deepfakes and the New Disinformation War: The Coming of Age of Post-Truth Geopolitics." *Foreign Affairs* 98, no. 1 (2019): 147–155.

8. Çiftçi, Umur A., Gokturk Yuksek and Ilke Demir. "My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymization." *arXiv* (Accessed 13 December, 2022) // https://arxiv.org/abs/2211.01361.

9. Coeckelbergh, Mark. *AI Ethics*. Cambridge (MA) and London: The MIT Press, 2020.

10. Constine, Josh. "ByteDance and TikTok Have Secretly Built a Deepfakes Maker." *TechCrunch* (January 2020) // https://techcrunch.com/2020/01/03/tiktok-deepfakes-face-swap/.

11. Cover, Rob. "Deepfake Culture: The Emergence of Audio-Video Deception as an Object of Social Anxiety and Regulation." *Continuum* 36, no. 4 (2022): 609–621.

12. De Souza, Siddharth Peter. "The Spread of Legal Tech Solutionism and the Need for Legal Design." *European Journal of Risk Regulation* 13, no. 3 (2022): 373–388.

13. Di Gregorio, Giovanni. *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society*. Cambridge and New York: Cambridge University Press, 2022.

14. Diakopoulos, Nicholas and Deborah Johnson. "Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections." *New Media & Society* 23, no. 7 (2021): 2072–2098.

15. Diresta, Renee. "AI-Generated Text Is the Scariest Deepfake of All." *Wired* (July 2020) // https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all/.

16. Du Sautoy, Marcus. *The Creativity Code: How AI Is Learning to Write, Paint and Think*. London: 4th Estate, 2020.

17. Edwards, Benj. "AI Image Generation Tech Can Now Create Life-Wrecking Deepfakes with Ease." *Ars Technica* (December 2022) // https://arstechnica.com/information-technology/2022/12/thanks-to-ai-its-probably-time-to-take-your-photos-off-the-internet/.

18. Fallis, Don. "The Epistemic Threat of Deepfakes." *Philosophy and Technology* 34 (2020): 223–243.

19. Feingold, Spencer. "Artificial Intelligence Image Generators Bring Delight – and Concern." *World Economic Forum* (October 2022) // https://www.weforum.org/agenda/2022/10/ai-artist-systems-bring-delight-and-concern/.

20. Galston, William A. "Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics." *Brookings* (January 2020) // https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/.

21. Gapper, John. "ChatGPT Is Fluent, Clever and Dangerously Creative." *Financial Times* (December 2022) // https://www.ft.com/content/86e64b4c-a754-47d6-999c-fcc54f62fb5d.

22. Giles, Keir, Kim Hartmann and Munira Mustaffa. *The Role of Deepfakes in Malign Influence Campaigns*. Riga: NATO StratCom COE, 2019.

23. Ging, Debbie. "Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere." *Men and Masculinities* 22, no. 4 (2019): 638–657.

24. Goldman, Sharon. "The Hidden Danger of ChatGPT and Generative AI." *VentureBeat* (December 2022) // https://venturebeat.com/ai/the-hidden-danger-of-chatgpt-and-generative-ai-the-ai-beat/.

25. Guadamuz, Andres. "Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works." In *Artificial Intelligence & Intellectual Property*, edited by Jyh-An Lee, Reto Hilty and Kung-Chung Liu. Oxford and New York: Oxford University Press, 2021, 147–176.

26. Harrington, Carol. "What is 'Toxic Masculinity' and Why Does it Matter?" *Men and Masculinities* 24, no. 2 (2021): 345–352.

27. Harsin, Jayson. "Aggro-Truth: (Dis-)Trust, Toxic Masculinity, and the Cultural Logic of Post-Truth Politics." *The Communication Review* 24, no. 2 (2021): 133–166.

28. Harsin, Jayson. "Toxic White Masculinity, Post-Truth Politics and the COVID-19 Infodemic." E*uropean Journal of Cultural Studies* 23, no. 6 (2020): 1060–1068.

29. Heilweil, Rebecca. "AI Is Finally Good at Stuff, and that's a Problem." *Vox* (December 2022) // https://www.vox.com/recode/2022/12/7/23498694/ai-artificial-intelligence-chat-gpt-openai.

30. Helmus, Todd C. "Artificial Intelligence, Deepfakes, and Disinformation: A Primer." *RAND Corporation* (July 2022) // https://www.rand.org/pubs/perspectives/PEA1043-1.html.

31. Hern, Alex. "AI Bot ChatGPT Stuns Academics with Essay-Writing Skills and Usability." *The Guardian* (December 2022) // https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability.

32. Heslep, Daniel G. and P. S. Berge. "Mapping Discord's Darkside: Distributed Hate Networks on Disboard." *New Media & Society*, online first // DOI: 10.1177/14614448211062548.

33. Honigberg, Bradley. "The Existential Threat of AI-Enhanced Disinformation Operations." *Just Security* (July 2022) // https://www.justsecurity.org/82246/the-existential-threat-of-ai-enhanced-disinformation-operations/.

34. Illia, Laura, Elanor Colleoni and Stelios Zyglidopoulos. "Ethical Implications of Text Generation in the Age of Artificial Intelligence." *Business Ethics, the Environment & Responsibility* 32, no. 1 (2022): 201–210.

35. Jacobs, Harrison. "Getty Images Bans AI-Generated Images due to Copyright Worries." *ARTnews* (September 2022) // https://www.artnews.com/art-news/news/getty-images-bans-ai-generated-images-due-to-copyright-1234640201/.

36. Kalpokas, Ignas. "Posthuman Urbanism: Datafication, Algorithmic Governance and Covid-19." In *The Routledge Handbook of Architecture, Urban Space and Politics Volume I: Violence, Spectacle and Data*. Edited by Nikolina Bobic and Farzaneh Haghighi. London and New York: Routledge, 2023, 496–508.

37. Kalpokas, Ignas. "Problematising Reality: The Promises and Perils of Synthetic Media." *SN Social Sciences* 1, no. 1 (2021b): 1–11.

38. Kalpokas, Ignas. *Algorithmic Governance: Law and Politics in a Post-Human Era*. Cham: Palgrave Macmillan, 2019.

39. Kalpokas, Ignas. *Malleable, Digital, and Posthuman: A Permanently Beta Life*. Bingley: Emerald Publishing Limited, 2021a.

40. Kietzmann, Jan, Linda W. Lee, Ian P. McCarthy, and Tim C. Kietzmann. "Deepfakes: Trick or Treat." *Business Horizons* 63, no. 2 (2020): 135–146.

41. Knight, Will. "This Copyright Lawsuit Could Shape the Future of Generative AI." *Wired* (November 2022) // https://www.wired.com/story/this-copyright-lawsuit-could-shape-the-future-of-generative-ai/.

42. Knight, Will. "This Uncensored AI Art Tool Can Generate Fantasies – and Nightmares." *Wired* (September 2022b) // https://www.wired.com/story/the-joy-and-dread-of-ai-image-generators-without-limits/.

43. Kreps, Sarah, McCain, R. Miles and Brundage, Miles. "All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation." *Journal of Experimental Political Science* 9 (2022): 104–117.

44. Lin, Connie. "How to Easily Trick OpenAI's Genius New ChatGPT." *Fact Company* (December 2022) // https://www.fastcompany.com/90819887/how-to-trick-openai-chat-gpt.

45. Loizos, Connie. "Is ChatGPT a 'Virus that Has Been Released into the Wild'?" *TechCrunch* (December 2022) // https://tcrn.ch/3W7IZUY.

46. Lovato, Juniper et al. "Diverse Misinformation: Impacts of Human Biases on Detection of Deepfakes on Networks." *arXiv* (Accessed 13 December, 2022) // https://arxiv.org/abs/2210.10026.

47. Marelli, Luca, Kieslich, Katharina and Geiger, Susi. "COVID-19 and Technosolutionism: Responsibilization without Contextualization?" *Critical Public Health* 32, no. 1 (2022): 1–4.

48. Massanari, Adrienne. "#Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures." *New Media & Society* 19, no. 3 (2015): 329–346.

49. McCosker, Anthony. "Making Sense of Deepfakes: Socializing AI and Building Data Literacy on GitHub and YouTube." *New Media & Society* (2022) // https://doi.org/10.1177/14614448221093943.

50. McGowran, Leigh. "'Legal Minefield': The Risk of Commercialising AI-Generated Images." *SiliconRepublic* (September 2022) // https://www.siliconrepublic.com/machines/ai-generated-images-legal-risks-copyright.

51. Metz, Cade. "The New Chatbots Could Change the World. Can You Trust Them?" *The New York Times* (December 2022) // https://www.nytimes.com/2022/12/10/technology/ai-chat-bot-chatgpt.html.

52. Metz, Rachel. "Anyone Can Now Use Powerful AI Tools to Make Images. What Could Possibly Go Wrong?" *CNN* (October 2022) // https://edition.cnn.com/2022/09/30/tech/image-generating-ai-publicly-available/index.html.

53. Miller, Arthur I. *The Artist in the Machine: The World of AI-Powered Creativity*. Cambridge (MA) and London: The MIT Press, 2019.

54. Morozov, Evgeny. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: Public Affairs, 2014.

55. Mullen, Molly. "New Reality: Deepfake Technology and the World Around Us." *Mitchell Hamline Law Review* 48, no. 1 (2022): 210–234.

56. Newton, Olivia B. and Mel Stanfill. "My NSFW Video Has Partial Occlusion: Deepfakes and the Technological Production of Nonconsensual Pornography." *Porn Studies* 7, no. 4 (2020): 398–414.

57. Nightingale, Sophie J. and Hany Farid. "AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy." *Proceedings of the National Academy of Sciences* 119, no. 8 (2022): 1–3.

58. Pawelec, Maria. "Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions." *Digital Society* 1 (2022): 1–37.

59. Potts, John. *The Near-Death of the Author: Creativity in the Internet Age*. Toronto: The University of Toronto Press, 2022.

60. Robinson, Rebekah. "AI Image Generators Enable the Creation of Fake Pictures to Support Fake News." *Coda* (October 2022) // https://www.codastory.com/disinformation/ai-image-generators-fake-news/.

61. Salter, Anastasia and Bridget Blodgett. *Toxic Geek Masculinity in Media: Sexism, Trolling, and Identity Policing*. New York: Palgrave Macmillan, 2017.

62. Simonite, Tom. "Forget Politics: For Now, Deepfakes are for Bullies." *Wired* (September 2019) //https://www.wired.com/story/forget-politics-deepfakes-bullies/.

63. Stokel-Walker, Chris. "AI Bot ChatGPT Writes Smart Essays — Should Professors Worry?" *Nature* (December 2022) // https://www.nature.com/articles/d41586-022-04397-7.

64. Sullivan, Mark. "Meta's New AI Video Generator Could Be a Dangerous Misinformation Tool." *FastCompany* (September 2022) // https://www.fastcompany.com/90793071/metas-new-ai-video-generator-could-be-a-dangerous-misinformation-tool.

65. Taylor, Linnet. "There Is an App for That: Technological Solutionism as COVID-19 Policy in the Global North." In *The New Common: How the COVID-19 Pandemic Is Transforming Society.* Edited by Emile Aarts et al. Cham: Springer, 2021, 209–215.

66. Ten Oever, Niels. "This Is not how We Imagined It: Technological Affordances, Economic Drivers, and the Internet Architecture Imaginary." *New Media & Society* 23, no. 2 (2021): 344–362.

67. Thomas, Elise. "In the Battle Against Deepfakes, AI Is Pitted against AI." *Wired* (November 2019) // https://www.wired.co.uk/article/deepfakes-ai.

68. Tolosana, Ruben et al. "An Introduction to Digital Face Manipulation." In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*. Edited by C. Rathgeb et al., 3–26. Cham: Springer, 2022.

69. Trott, Verity Anne. "Gillette: The Best a Beta Can Get. Networking Hegemonic Masculinity in the Digital Sphere." *New Media & Society* 24, no. 6 (2022): 1417–1434.

70. Vaccari, Christian and Chadwick, Andrew. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media + Society* 6. No. 1 (January-March 2020) // https://doi.org/10.1177/2056305120903408.

71. van der Nagel, Emily. "Verifying Images: Deepfakes, Control, and Consent." *Porn Studies* 7, no. 4 (2020): 424–429.

72. van der Sloot, Bart and Yvette Wagensveld. "Deepfakes: Regulatory Challenges for the Synthetic Society." *Computer Law & Security Review* 46 (2022): 1–15.

73. Vincent, James. "Getty Images Bans AI-Generated Content over Fears of Legal Challenges." *The Verge* (September 2022b) // https://www.theverge.com/2022/9/21/23364696/getty-images-ai-ban-generated-artwork-illustration-copyright.

74. Vincent, James. "The Lawsuit that Could Rewrite the Rules of AI Copyright." *The Verge* (November 2022) // https://www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data.

75. Vizoso, Ángel, Martín Vaz-Álvarez and Xosé López-García. "Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation." *Media and Communication* 9, no. 1 (2022): 291-300.

76. Waling, Andrea. "Problematising 'Toxic' and 'Healthy' Masculinity for Addressing Gender Inequalities." *Australian Feminist Studies* 34, no. 101 (2019): 362–375.

77. Whittaker, Lucas, et al. "All Around Me Are Synthetic Faces: The Mad World of AI-Generated Media." *IT Professional* 22, no. 5 (2020): 90–99.

78. Woolley, Samuel. *The Reality Game: How the Next Wave of Technology Will Break the Truth and What We Can Do About It*. London: Endeavour, 2020.

79. Zeilinger, Martin. *Tactical Entanglements: AI Art, Creative Agency, and the Limits of Intellectual Property*. Lüneburg: Meson Press, 2021.