California Management Review

TECHNOLOGY

Synthetic Media: What Managers Need to Know about this Emergent Phenomenon

by Lucas Whittaker, Andrew Park, and Jan Kietzmann

Image Credit | Provided by Lucas Whittaker from the 2018 "J'adore Dior the New Absolu" commercial

Recommendations to prepare for and mitigate potential harmful effects of synthetic media incidents.

INSIGHT | FRONTIER 17 Oct 2022

Synthetic media, comprised of manipulated digital content such as deepfake videos, audio, and images, have garnered increased public attention recently. They have been used in creative and productive ways to achieve organizational goals, as seen in the wildly successful social media campaign of South Korean presidential candidate, Yoon Suk-yeol, whose team generated numerous deepfake memes of the candidate to appeal to younger voters.[1] However, synthetic media have also facilitated the actions of malicious actors. In the case of the "Fake Johannes", fraudsters cloned the voice of the CEO of a German energy firm using deepfake audio, and through a telephone call, convinced an executive at a subsidiary company to wire US$231,000 to the fraudsters' own account at a Hungarian bank.[2]

RELATED CMR ARTICLES

"Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing" by V. Kumar, Bharath Rajan, Rajkumar Venkatesan, & Jim Lecinski. (Vol. 61/4) 2019.

Realistic and convincing synthetic media are largely enabled by rapid developments in artificial intelligence and popular machine learning methods, in particular deep learning. As these technologies continue to evolve, it is expected that an increasing number of brands, governments, and other organizations will be susceptible to synthetic media attacks.[3] The harmful outcomes of these attacks include loss of financial resources and critical assets, erosion of brand reputation, loss of customer and end-user confidence, and damage to an organization's security infrastructure. Thus, it is becoming increasingly important that managers become familiar with malicious synthetic media and devise a plan to protect their organizations against their emergence

In this article, we provide a short overview of the evolution of media manipulation. We then comment on why synthetic media in particular are so convincing to humans and why they present a unique threat to organizations today. We conclude by offering several

managerial recommendations to prepare for and mitigate the potential harmful effects of synthetic media incidents.

The state of digital media manipulation

At its core, *digital media manipulation* uses computers to alter existing media, or produce altogether new media.[4] The ability for the average consumer to participate in the production of manipulated images is largely credited to the introduction of Adobe Photoshop in the 1990s. Today, manipulated images, audio clips, and videos can be created and disseminated by a multitude of non-technical audiences, even via mobile phones through easy-to-use filters on popular social media applications such as Snapchat and Instagram.

The more recent genesis of *synthetic media manipulation* can arguably be traced back to a Reddit thread in 2017 when a user by the name of 'Deepfakes' posted a series of synthetically created videos of celebrities in compromising and salacious settings. These videos were created using face-swapping technology to splice existing photos of celebrities and digitally graft their faces onto actors in the questionable videos.

In contrast to the process of 'photoshopping', this technology leverages a subset of machine learning termed deep learning - a cyclical computer algorithm that iteratively improves its ability to identify commonalities between the donor's and recipient's faces (or voices) and ultimately replace the recipient's features with that of the donor's.[5] Through such synthetic manipulation, one can automatically, not manually, replace the faces of actors in an advertisement, for instance (as shown in Figure 1 below). GANs can take this algorithmic manipulation even further, not merely swapping content, but creating new content altogether.[6] As examples, GANs have been used to create original artwork in the style of Monet and Van Gogh,[7] and even entire symphonies and depictions of actual urban environments.[8] In short, deepfakes and GANs present compelling, customized, and

automatically fabricated images, audio, or videos featuring people or objects in settings that never occurred. When done well, the results are impossible to tell apart from authentic content.

Because of the democratization of synthetic media manipulation, i.e., that it requires less and less technical skill for a user to generate increasingly convincing content, it is expected that synthetic media will pose a growing threat to individuals and organizations. A significant contributor to the seriousness of this threat is the phenomenon known as *nonveridicality*[9] – media may inaccurately or inauthentically portray phenomena; however, the perceiver perceives the information as legitimate. Humans have always relied on cognitive shortcuts to tell apart what they perceive to be true and genuine from what they believe is fake. We have long relied on media to signal such authenticity, especially multimodal media content that closely emulates the real world. Such media are seen as credible. One only needs to think about photos and videos used as evidence in a court of law. We are primed to trust our eyes – to believe what we see.

Against this backdrop, it is easy to see how synthetic media create nonveridical perceptions. The broad accessibility of consumer-friendly apps such as Snapchat and Instagram, and the availability of increasingly user-friendly tools that facilitate the automatic development of highly vivid and realistic-looking content should worry managers. All organizations and industries are increasingly vulnerable to a host of threats introduced through synthetic media manipulation.

Synthetic media threats to organizations

We observe several key threats related to media manipulation facing managers today. First, there exists significant reputational risk if an organization is poorly equipped to defend itself against a synthetic media attack. When an organization loses control of its reputation, it can lead to rapid and material erosion of brand trust, particularly if the damaging synthetic media are targeted towards generating broad mistrust of a brand's

values. This risk also extends to individuals within organizations, as evidenced by the recent circulation of a lewd video clip, purported to be of Malaysia's Minister of Economic Affairs, Mohamed Azmin Ali.

Second, managers and organizations face economic risks associated with the rise of synthetic media manipulation. While the aforementioned case of the "Fake Johannes" is likely the most well-known case of financial loss due to a deepfake attack, it is noteworthy that since then, similar incidents have been occurring consistently across numerous industries and organizations. Researchers at Symantec identified at least three cases of cyber-impersonation campaigns similar in style to that of the "Fake Johannes" in just several months after the infamous attack.[10] Even seemingly harmless community-driven endeavors such as the recent proliferation of deepfake versions of a Dior commercial which originally featured Charlize Theron (**Figure 1**), is broadly indicative of the potential loss of control and desired messaging for brands and organizations.

Figure 1. Deepfakes of Charlize Theron, sourced from the 2018 J'adore Dior the New Absolu commercial[11]

Third, managers must contend with intellectual property risks if there is a synthetic media-based organizational breach. In addition to trade secrets, technical know-how, and formalized intellectual property (such as patents) being compromised through a sophisticated infiltration event using deepfakes, managers must be acutely aware of the risk of trademark infringement, particularly because of the wide accessibility of tools that allow for the generation of deepfakes. Section 230 of the Communications Decency Act provides protection to social media companies by effectively removing any liability caused by the content created by its users. However, this immunity from civil liabilities becomes murky when applied to trademark infringement. In 2019, a deepfake of Mark Zuckerberg generated using source material from CBS began circulating widely on Facebook. This video retained the CBSN logo, and CBS promptly demanded that Facebook take down the video. Similarly, in the deepfakes shown above, despite possibly appearing harmless, the

brands of Dior, Theron, Robbie, Jolie, and Atkinson were used without their consent. Such incidences suggest that organizations may need to litigate if they fall prey to deepfakes that compromise their brands or intellectual property portfolio, which will require increasing the time and resources dedicated to their legal strategy.

Managerial recommendations

The preceding discussion on the looming risks of synthetic media are alarming due to the increasing believability and quantity of such media. This emergence is due, in part, to the broad and growing availability of easy-to-use, consumer-grade apps that present low barriers to entry in creating deepfakes5. We conclude with three recommendations for managers whose industries may be disrupted by synthetic media, which we adapt from the 2021 Cybersecurity Incident & Vulnerability Response Playbooks, issued by the US Cybersecurity & Infrastructure Security Agency.[12]

First, managers who prepare now will be in a better position to defend their organizations against a potential synthetic media incident. As an important first step, they can educate their team members on what synthetic media manipulation is and provide them some guidance on how to determine whether a video or audio file is likely to have been manipulated. Managers may elect to first focus on team members who have fiduciary privileges, or access to important trade secrets. Training efforts could then be extended to the rest of the team or even external stakeholders such as key customers so that they are also aware of the current threat environment.

Second, managers should identify and prioritize core assets when developing their cybersecurity response strategy. Relatedly, organizations can identify and create profiles of theoretical malicious parties that may want to appropriate or damage these assets. If the malicious actor is likely to be a technologist who might stage a sophisticated attack on the firm, managers can recruit technical employees to pre-emptively improve the

cybersecurity infrastructure. If the malicious actor is non-technical, and likely to be more interested in damaging the organization's brand reputation, the organization may instead focus its defense efforts on bolstering its brand image and messaging.

Third, if an organization has already suffered a synthetic media attack, it should broaden its net of security partners to prevent future incidents. Managers can alert jurisdictional police authorities and regulators, such as local securities commissions. Such authorities can monitor the local market for an uptick in similar incidences and create new social infrastructure to prevent future incidents, such as the deployment of educational campaigns to help its stakeholders identify common features of synthetic media attacks.

If alerted to a security incident, governmental authorities could potentially provide valuable protections and be ready to mobilize quickly if an attack recurs. However, given the increasing quantity and availability of deepfake generation tools, there is little guarantee that future incidents will not continue to occur. Thus, managers would benefit by acting now – helping their teams become familiar with synthetic media and understanding their risks, identifying their core assets, and building relationships with local authorities who may become valuable partners in the response to this emergent synthetic risk.

References

1. France24. (2022). Deepfake democracy: South Korean candidate goes virtual for votes. France 24. https://www.france24.com/en/live-news/20220214-deepfake-democracy-south-korean-candidate-goes-virtual-for-votes

2. C. Stupp. (2019, August 30). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. The Wall Street Journal. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

3. M. E. Bonfanti. (2020, July 14). The weaponisation of synthetic media: what threat does this pose to national security? https://www.realinstitutoelcano.org/en/analyses/the-weaponisation-of-synthetic-media-what-threat-does-this-pose-to-national-security/

4. C. Campbell, K. Plangger, S. Sands., & J. Kietzmann. (2021). Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. Journal of Advertising, 51(1), 22-38.

5. For more details on this process, see J. Kietzmann, L. Lee, I. McCarthy, & T.C. Kietzmann. (2020). Deepfakes: Trick or treat?. Business Horizons, 63(2), 135-146.

6. (#_ednref6) L. Whittaker, T.C. Kietzmann, J. Kietzmann, & A. Dabirian. (2020). All around me are synthetic faces: the mad world of AI-generated media. IT Professional, 22(5), 90-99.

7. (#_ednref7) J. Zhu, T. Park, P. Isola, & A. Efros. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2223-2232).

8. (#_ednref8) R. Gadde, Q. Feng, & A. M. Martinez. (2021). Detail Me More: Improving GAN's Photo-Realism of Complex Scenes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 13950-13959).

9. H. Lee, S. A. Park, Y. Lee, & G. T. Cameron. (2010). Assessment of motion media on believability and credibility: An exploratory study. Public Relations Review, 36(3), 310-312.

10. D. Harwell. (2019, September 04). An artificial-intelligence first: Voice-mimicking software reportedly used in a major theft. The Washington Post. https://www.washingtonpost.com/technology/2019/09/04/an-artificial-intelligence-first-voice-mimicking-software-reportedly-used-major-theft/

11. For a more convincing version of this manipulation, view the video at https://bit.ly/jadoredfs

12. Whittaker, J. Kietzmann, K. Letheren, R. Mulcahy, & R. Russell-Bennett. (2022). Brace yourself! Why Managers should adopt a Synthetic Media Incident Response Playbook in an age of Falsity and Synthetic Media. Business Horizons, forthcoming.

Lucas Whittaker Follow

Lucus Whittaker is a PhD Candidate at the Centre for Behavioural Economics, Society and Technology (BEST) at the Queensland University of Technology. His research focuses on the application of synthetic media such as ‘deepfakes’ within marketing contexts and psychological factors which may influence consumer appraisal of such media.

Andrew Park

Andrew Park, PhD is an Assistant Professor of Information Systems at the Gustavson School of Business, University of Victoria, Canada. His work centers around how Open Innovation mechanisms impact value creation of firms within the emerging Personalized Medicine ecosystem.

Jan Kietzmann

Jan Kietzmann, PhD is a Professor of Innovation and Information Systems at the Gustavson School of Business, University of Victoria, Canada. The focus of Jan’s work is on organizational and social perspectives related to innovation and emerging technologies.