CURRENT ISSUES IN TOURISM
2021, VOL. 24, NO. 13, 1798–1802
https://doi.org/10.1080/13683500.2020.1738357

Routledge
Taylor & Francis Group

RESEARCH LETTER

Deepfake: a social construction of technology perspective

Andrei O. J. Kwok [a,b] and Sharon G. M. Koh[c]

[a]Department of Management, Sunway University Business School, Sunway University, Malaysia; [b]Department of Management, School of Business, Monash University, Malaysia; [c]Department of Economics, School of Business, Monash University, Malaysia

ABSTRACT

The advent of deepfake videos, a new form of media manipulation, took the world by storm. Malicious use of this technology for fraud, misrepresentation, and abuse created more perils than opportunities for various stakeholders. However, little is known about Generative Adversarial Networks utilized by deepfake to manipulate videos. Viewing from the social construction of technology theory, we explore the redress of the adversarial purposes of deepfake. At the moment, there are more misgivings than credence about the possible application of this technology. Our research note provides a discourse on the rapid advances of deepfake and its potential impact on tourism and raises the call for research on the beneficial uses of deepfake.

ARTICLE HISTORY
Received 12 January 2020
Accepted 28 February 2020

KEYWORDS
Artificial intelligence; video/imaging technology; generative adversarial networks; destination image; tourism marketing

Introduction

Recent evidence indicates that the propagation of false information can mislead and potentially trigger severe complications in society (Fedeli, 2020; Zhang & Ghorbani, 2020), exemplified in an altered photo of Bali's Gates of Heaven posted by Instagram influencers (Swaragita, 2019). The advancement of digital imaging using artificial intelligence (AI) and machine learning in tourism has given rise to immersive but manipulatable experiences such as extended reality (virtual, augmented, and mixed) that engage users in cyber and physical interactions (Sigala, 2018). However, unlike extended reality that involves human-machine interaction, deepfake could 'merge, combine, replace and superimpose' various forms of media to easily produce synthetic media that obscure the distinction between real and fake information (Maras & Alexandrou, 2019, p. 255). Hence deepfake tends to exploit the social media platforms as a means to spread false information.

Used maliciously, the proliferation of deepfake videos circulating online is a rising cause for concern in the tourism industry, especially the potential impact on the destination image and branding since detection is relatively challenging. While the increase of fake news is alarming, *'videos appeal at a visceral level more than any text or picture ever will, and the danger of the hoax spreading and shaping our perceptions is many times higher'* (Chakhoyan, 2018).

Given that the impact and the use of deepfake have been relatively unexplored, especially in tourism (Westerlund, 2019), our research note offers an alternative theory of formalization to explain the relation between deepfake and the real world, using Pinch and Bijker's (1984) social construction of technology theory. Despite the profusion about the malicious use of deepfake in the media, are there possible benefits for deepfake to have a positive impact on tourism? With the increasing use of online space, is there a preventive mechanism to govern deepfake's application or improve its detection? As such, we attempt to advance the research agenda via three overarching propositions.

CONTACT Andrei O. J. Kwok andreik@sunway.edu.my

Generative adversarial networks

Deepfake utilizes Generative Adversarial Networks (GANs) to manipulate all forms of media (photographs, texts, audio, and videos). GANs is a machine learning innovation that simultaneously trains two competing models (neural networks), comprising a generator and a discriminator to synthesize images (Goodfellow et al., 2014). The algorithms mirror the content and produce realistic but fake images and videos. The process involved in generating deepfake is shown in Figure 1.

Initially, the generator randomly draws data from the latent space and attempts to produce fake images that mimic real images input from the training dataset. The discriminator will then attempt to differentiate between real and fake images. As both models compete, improvements are made simultaneously, ending with convergence or equilibrium, whereby the generator will improve to the point when the discriminator increasingly faces difficulty in differentiating between real and fake images (see Brock et al., 2018).

Social construction of technology

The social construction of technology theory developed by Pinch and Bijker (1984) and extended to the digital world by van Baalen et al. (2016) views users as technological change agents. Multiple groups of users, as the stakeholders who adopt, apply and share the meanings of the technology define the trajectory of the technology development via a construction process of the embodiment of interpretation of its artifact and negotiation of its designs (Klein & Kleinman, 2002). The basic premise of this research note is not about rationalizing whether deepfake/GANs is a 'good' or 'bad' technology, but to debate the potential positive application and regulation of deepfake. The intent is also to call researchers to consider what beneficial uses of the technology that can be developed to preponderate over the existing negative aspects.

Potential impact and application in tourism

Proposition 1: novelty effect will lead to increased deployment of deepfake in the tourism industry

The advancement of deepfake as a media enhancing technology presents immense potential in various industries, e.g. healthcare, and entertainment. Especially in healthcare, deepfake has proven to offer significant diagnostic benefits (see Rees, 2019; and Westerlund, 2019). In tourism, the photorealistic images of scenes generated through this technology are especially useful in

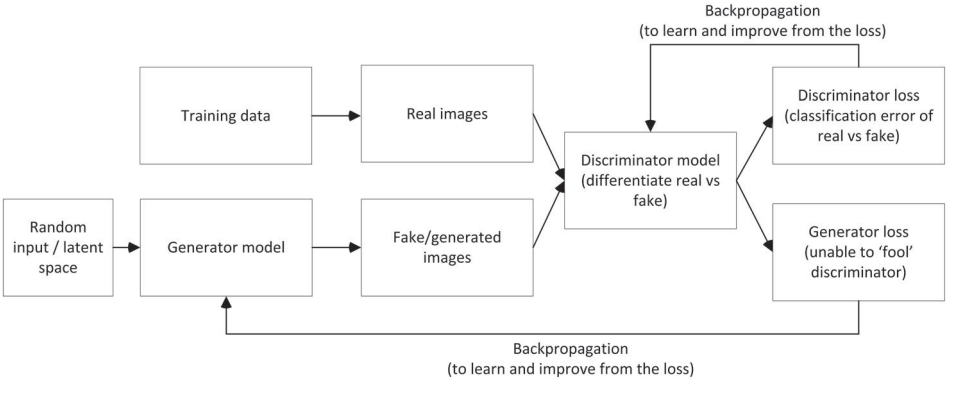

Figure 1. Generative Adversarial Network architecture. Source: Authors' adaptation from developers.google.com/machine-learning/gan.

advertising and marketing content creation (Picazo & Moreno-Gil, 2019). Images can offer a destination ad-evoked positive affect on tourists (Hosany et al., 2019) and offer companies user co-innovation opportunities (Chen et al., 2019). Furthermore, super-resolution media images can create a positive impression whereby tourists can be included to co-create the destination image using the social media platform (Cimbaljević et al., 2019).

Proposition 2: deepfake will improve user experience and enhance marketing

Deepfake can enhance visitor learning engagement and experience in galleries. Museums worldwide can benefit from this revolutionary convergence between art and artificial intelligence, thus bringing famous artists and their artworks to life. Such experience is possible through the Dalí Lives exhibition in the Dalí museum, St. Petersburg, USA. The museum uses deepfake to recreate an immersive visitor interaction and learning about Salvador Dalí (1904-1989) from the renowned artist himself (Lee, 2019). Improved deepfake techniques, which overlay high-resolution image blending (Wu et al., 2019) and image-to-image translation (Isola et al., 2017), can transform visitors' experience, thus adding a new dimension of personalized service. With the demand for celebrity images in advertisements and destination marketing (Van der Veen & Song, 2014), digitally realistic rendition of videos without the celebrity having actually to perform in it has proven possible with significant flexibility (Rees, 2019). Another example that could be adopted in tourism promotion is Québec's Crown corporation's 50th-anniversary advertisement using deepfake to reminisce viewers of a 1970s version of Bernard Derome, a well-known Canadian news anchor (La Presse Canadienne, 2020).

Proposition 3: concerns for deepfake misappropriation will improve governance and authentication mechanisms

As deepfake progresses in capability and becomes easily accessible, the question that arises is whether *there is a preventive mechanism to govern its applications or improve authentication?* It is relatively easy for the upsurge of doctored photos and videos with little detectable evidence of manipulation to make their rounds on social media. Deepfake media of corporate and public figures could result in individuals suffering from reputation damage, with possible extension to entities for which celebrities act as brand ambassadors (Eadicicco, 2019). Moreover, a rare first case of cybercrime using voice mimicry has been reported (Stupp, 2019).

As tourists tend to rely on social media imageries, and often, instead of determining the authenticity themselves, they too may reflexively circulate the imageries (Jansson, 2018). The inability to distinguish between real and fake will lead to misrepresentation, confusion, and erosion of trust about a destination. Confidence in visit decisions can be undermined or misplaced. Although at the moment, no destination, in particular, is known to have suffered from any deepfake harms yet, its rapid advancement could eventually lead to other unprecedented risks such as security and privacy breach that are not confined to image and branding. The potential harm will prompt providers and government regulators to step up investment into investigation tools to safeguard the rights of those affected (Brundage et al., 2018).

Current conundrum

While nascent, detection and regulation are at work. For example, California has issued temporary and specific bans of political deepfakes from influencing an election, whereas, in China, the Cyberspace Administration of China, the Ministry of Culture and Tourism, and the National Radio and Television Administration have jointly banned the creation, broadcast, and use of deepfake (Jing, 2019). However, given the many existing loopholes, banning deepfakes may seem to raise more questions instead, as the more underlying problem is the spread of disinformation (Morris, 2020). Meanwhile, the Media Forensics under the United States Defense Advanced Research Projects Agency (DARPA) is

funding technologies for automated detection and assessment of visual media integrity and facilitates investigation of the manipulation performed[1] (Maras & Alexandrou, 2019). Technology companies such as Facebook, Twitter, and Google are imposing limited bans of deepfakes (Morris, 2020), while also combating deepfakes by releasing database of deepfakes to improve identification (Rees, 2019). Despite such efforts, the conundrum is whether the improvement rate of detection can outpace the rapid advances of deepfake capability.

Conclusion

This research note explores the emergence of GANs that has been coopted for adversarial purposes in the form of deepfake that potentially impacts different tourism scholars, practitioners, and researchers. Given the infancy of its real-word application, there are still many unanswered questions on the potential impact of deepfake. We debated the impact and application from the lens of the social construction of technology theory, in that, while it is difficult to arrest the march of technology, the potential use of such innovation offers significant beneficial opportunities. Since our paper is one of the earliest to explore the use of deepfake in tourism, sources of references are mostly technical papers and online news due to a lack of scholarly studies in social science. We encourage future studies to expand on real-word case studies of deepfake application and to examine its effects empirically. Researchers are also encouraged to consider the ethics of deepfake in tourism marketing. Theoretically, studies could also view from Feenberg's (2002) critical theory, where technology could be deliberated not as an object but based on the value it brings under different social contexts.

Note

1. See DARPA (www.darpa.mil/program/media-forensics).

Acknowledgements

The authors are grateful to the editor and anonymous reviewers for their valuable feedback in improving the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Andrei O. J. Kwok http://orcid.org/0000-0002-2733-9931

References

Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. *arXiv*. https://arxiv.org/abs/1809.11096

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., & Anderson, H. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv*. https://arxiv.org/abs/1802.07228

Chakhoyan, A. (2018, November 16). Deep fakes could threaten democracy. What are they and what can be done? *World Economic Forum*. https://www.weforum.org/agenda/2018/11/deep-fakes-may-destroy-democracy-can-they-be-stopped/

Chen, S., Kang, J., Liu, S., & Sun, Y. (2019). Cognitive computing on unstructured data for customer co-innovation. *European Journal of Marketing*. https://doi.org/10.1108/EJM-01-2019-0092

Cimbaljević, M., Stankov, U., & Pavluković, V. (2019). Going beyond the traditional destination competitiveness–reflections on a smart destination in the current research. *Current Issues in Tourism*, *22*(20), 2472–2477. https://doi.org/10.1080/13683500.2018.1529149

Eadicicco, L. (2019, July 10). There's a terrifying trend on the internet that could be used to ruin your reputation, and no one knows how to stop it. *Business Insider US*. https://www.businessinsider.my/dangerous-deepfake-technology-spreading-cannot-be-stopped-2019-7/?r=US&IR=T

Fedeli, G. (2020). 'Fake news' meets tourism: A proposed research agenda. *Annals of Tourism Research*, *80*, 102684. https://doi.org/10.1016/j.annals.2019.02.002

Feenberg, A. (2002). *Transforming technology: A critical theory revisited* (2nd ed.). New York, N.Y.: Oxford University Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). Generative adversarial nets. *arXiv*. https://arxiv.org/abs/1406.2661

Hosany, S., Buzova, D., & Sanz-Blas, S. (2019). The influence of place attachment, Ad-evoked positive affect, and motivation on intention to visit: Imagination proclivity as a moderator. *Journal of Travel Research*, *59*(3), 477–495. https://doi.org/10.1177/0047287519830789

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *arXiv*. https://arxiv.org/abs/1611.07004

Jansson, A. (2018). Rethinking post-tourism in the age of social media. *Annals of Tourism Research*, *69*, 101–110. https://doi.org/10.1016/j.annals.2018.01.005

Jing, M. (2019, November 29). China issues new rules to clamp down on deepfake technologies used to create and broadcast fake news. *South China Morning Post*. https://www.scmp.com/tech/apps-social/article/3039978/china-issues-new-rules-clamp-down-deepfake-technologies-used

Klein, H. K., & Kleinman, D. L. (2002). The social construction of technology: Structural considerations. *Science, Technology, & Human Values*, *27*(1), 28–52. https://doi.org/10.1177/016224390202700102

La Presse canadienne. (2020, February 11). Loto-Québec, Bernard Derome et la banalisation de l'hypertrucage. *Radio-Canada*. https://ici.radio-canada.ca/nouvelle/1514856/deepfake-bernard-derome-loto-quebec-risque-derape-consentement-jeff-yates-sid-lee-decrypteurs

Lee, D. (2019, May 10). Deepfake Salvador Dalí takes selfies with museum visitors. *The Verge*. https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226

Morris, B. (2020, January 7). Facebook bans deepfakes but permits some altered content. *The Wall Street Journal*. https://www.wsj.com/articles/facebooks-deepfake-video-ban-permits-some-altered-content-11578384519

Picazo, P., & Moreno-Gil, S. (2019). Analysis of the projected image of tourism destinations on photographs: A literature review to prepare for the future. *Journal of Vacation Marketing*, *25*(1), 3–24. https://doi.org/10.1177/1356766717736350

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Rees, G. (2019, November 25). *Here's how deepfake technology can actually be a good thing*. https://www.weforum.org/agenda/2019/11/advantages-of-artificial-intelligence/

Sigala, M. (2018). New technologies in tourism: From multi-disciplinary to anti-disciplinary advances and trajectories. *Tourism Management Perspectives*, *25*, 151–155. https://doi.org/10.1016/j.tmp.2017.12.003

Stupp, C. (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

Swaragita, G. (2019, July 6). Bali's Lempuyang temple 'gates of heaven' less dramatic than expected. *The Jakarta Post*. https://www.thejakartapost.com/travel/2019/07/06/balis-lempuyang-temple-gates-of-heaven-less-dramatic-than-expected.html

van Baalen, P. J., van Fenema, P. C., & Loebbecke, C. (2016). *Extending the social construction of technology (SCOT) framework to the digital world*. In ICIS.

Van der Veen, R., & Song, H. (2014). Impact of the perceived image of celebrity endorsers on tourists' intentions to visit. *Journal of Travel Research*, *53*(2), 211–224. https://doi.org/10.1177/0047287513496473

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. https://doi.org/10.22215/timreview/1282

Wu, H., Zheng, S., Zhang, J., & Huang, K. (2019). Gp-gan: Towards realistic high-resolution image blending. *arXiv*. https://arxiv.org/abs/1703.07195

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025. https://doi.org/10.1016/j.ipm.2019.03.004