

Is any deepfake usage for political purposes unethical? The case study of a Korean presidential candidate

Научный руководитель – Barry Marguerite

Lavrov Ilya

Студент (магистр)

Университетский Колледж Дублина, Дублин, Ирландия

E-mail: ilyalavrov0607@gmail.com

Traditionally, the dissemination of “AI-powered audio, image and video synthesis - so-called deepfakes” [9] is perceived as ethically controversial as it has potential to be harmful both to individuals and to society at large [2]. There are many concerns in academic literature about potential destructive implications of this technology in various fields [2, 5, 6, 8] as it possesses “unprecedented opportunities for deception”[3]. The political sphere is usually considered as one of the most vulnerable to deepfakes and is even described as a “a significant threat to democracy, national security, and society” [1]. When it comes to election campaigns, deepfakes can potentially be detrimental to 1) voters as a source of deception and intimidation; 2) to candidates and campaigns when posing reputational harm and resulting in misattribution; 3) and to social institutions, undermining trust in democratic elections [4].

Yoon Suk-yeol, a South Korean candidate, decided to use AI technologies to manipulate his own video samples to “create an impressively convincing digitalized image” which would be more appealing to voters [10]. However, he must have faced a dilemma: is it ethical or unethical to manipulate your own audio and video with the use of machine learning for political purposes when you are a presidential candidate? When it comes to manipulating a video sample of your opponent, it is usually considered unethical if it has been done without the consent of that person. However, Yoon Suk-yeol decided that it would be appropriate and even would make him more appealing to voters.

Is any deepfake unethical? Even if it is portrayed like this in social media, it cannot be stated that it is. To characterize a manipulated video as ethically problematic several factors should be taken into account: “(i) whether the deepfaked person(s) would object to the way in which they are represented; (ii) whether the deepfake deceives viewers; and (iii) the intent with which the deepfake was created” [3]. Using this theoretical framework: (i) Yoon Suk-yeol is willing to be portrayed in the way he is portrayed in those deepfake videos as he looks much more appealing on them than in reality, especially for the younger generation; (ii) the candidate openly states that he uses deepfakes during his election campaign and it means that people are aware of the fact that they are watching deepfakes, hence, the deceiving effect is minimized; (iii) those deepfakes are not made with malicious intent as those videos are produced to make a candidate more appealing to the public, to produce as many videos as possible, answering all of the electorate’s questions, according to the party’s political program. The intent is not to disseminate misinformation and fake news. Thus, using the criteria, suggested in [3], the avatar of Yoon Suk-yeol, which appeared on the campaign videos, should be considered morally acceptable.

Manipulated videos can be seen as an ethical technology when 1) moral norms are not violated or such violation is minimized (following deontology); 2) there are more positive consequences of the usage of machine learning technics for video making in this case than direct negative effects (following consequentialism); 3) creating and watching deepfake videos people’s characters and “virtue” are not weakened (following virtue ethics); 4) trust and relations between

people are not negatively affected, that usually depends on the context in which a deepfake was created and disseminated (following care ethics). Thus, for the chosen case, following the theoretical framework described above as well as the conceptualization of ethical and unethical technologies provided in [3], it can be considered ethical if a candidate creates deepfakes only with their identity and openly tells people about it (1), there is no severe harm caused by deepfakes and a candidate has more chances to win the race, voters are more informed about his program, a tech company works with a brand-new project and develops the algorithms and the only negative consequence is that a political opponent is more likely to lose the election (2); a candidate and his team have high moral values (3); the relations between people become better than before, thanks to deepfake videos, because, for instance, voters, who had not known the politician beforehand, found out about him, his party initiatives and his values (4).

To sum it all up, the case of Yoon Suk-yeol and his avatar, created with deep learning technologies proves that deepfakes are “morally suspect, but not inherently morally wrong” [3]. This case study indicates a milestone in political communication, and it symbolizes not only a solid shift in political communication strategies of Korean politicians, but also has potential to change the future of political campaigns all around the world.

Источники и литература

- 1) Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H., & Farid, H. (2019). Protecting World Leaders Against Deep Fakes. CVPR Workshops.
- 2) Chesney, R., & Citron, D. K. (2018). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. SSRN Electronic Journal.
- 3) de Ruiter, A. (2021). The Distinct Wrong of Deepfakes. *Philosophy & Technology*, 34(4), 1311–1332.
- 4) Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, 23(7), 2072–2098.
- 5) Fallis, D. (2021). The Epistemic Threat of Deepfakes. *Philosophy & Technology*, 34(4), 623–643.
- 6) Fletcher, J. (2018). Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. *Theatre Journal*, 70(4), 455–471.
- 7) Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2019). Transparency, Protest and Democratic Stability. *British Journal of Political Science*, 49(4), 1251–1277.
- 8) 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.
- 9) Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, 119(8), e2120481119.
- 10) Parry, R. L. (2022, February 15). Wannabe president is everywhere — thanks to his avatar. *World | The Times*. <https://www.thetimes.co.uk/article/korea-s-ai-yoon-avatar-stands-in-for-presidential-candidate-8kfr5r7bj>