Research on the Effects of "DeepFake" Technology for the Modern Digital Space

T. Afanasyeva and I. Yumasheva

Introduction

In recent years, fake news has become a serious menace to society. Fake news refers to content in the style of fictional news, fabricated to deceive the public. The dissemination of false information is carried out primarily through social networks, which makes it possible to achieve a negative impact on millions of users in a short time (Afanasyeva et al., 2019).

Currently, every fifth Internet user receives news through the online platform YouTube, followed by Telegram, Facebook, VKontakte and others. The growing popularity of video platforms underscores the need to create and develop tools to confirm the authenticity of news content. Given the ease of obtaining and spreading disinformation through social networks, online and video platforms, it is becoming increasingly difficult for users to determine the authenticity of incoming information content, which leads to negative consequences of the development of modern digital society and problems in the media (Mirzajonov, 2021).

Modern online users live in an era called the "post-truth era", characterized by digital disinformation, information wars, false information campaigns to manipulate public opinion (Borchard, 2018). Recent advances in information technology have provided conditions for the emergence of the DeepFake phenomenon, as a result of which hyperrealistic videos are created with the exchange of persons leaving a small number of traces of manipulation, completely invisible to users of video platforms and social networks. DeepFake is a product of artificial intelligence applications that

T. Afanasyeva (✉) · I. Yumasheva
St. Petersburg University of Management Technologies and Economics, St. Petersburg, Russian Federation
e-mail: i.yumasheva@spbacu.ru

A. Rumyantseva et al. (eds.), *Challenges and Solutions in the Digital Economy and Finance*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-14410-3_7

allow combining, replacing, and overlapping images and video clips to create fake videos that look authentic (Kambur, 2021).

DeepFake technology can generate, for example, a humorous, pornographic, or political video of a person when his image and voice are fully used. The scale of the spread of DeepFake and the availability of technology indicates that almost anyone with the skills of a confident user of computer technology can fabricate fake videos that are practically indistinguishable from the original media. Early examples of DeepFake focused on political leaders, actresses, comedians, and other media personalities. At the present stage of development, DeepFake technologies are used not only to attract attention and revenge, but also, in court as evidence for trial, political sabotage, terrorist propaganda, blackmail, etc. (Yumasheva & Grigorieva, 2021).

In order to understand the very essence of the creation technology, it is important to note that DeepFake relies on neural information networks that analyze large sets of data samples, while completely imitating a person's facial expression, mannerisms, voice and intonation. The process of creating DeepFake involves uploading videos of two people into a digital "deep learning" algorithm to teach it to change faces. In other words, DeepFake uses 3D face mapping technology, which is an audio visualization and 3D projection on a physical object of the environment, considering its geometry and location in space. 3D-mapping technologies and artificial intelligence allow you to change a person's face in a video to another person's face, as a result, a complete substitution of the personality as the object of the video is made (Bondarchuk & Kotliar, 2020).

The purpose of the study is to develop recommendations on methods to combat DeepFake technologies. Before formulating effective methods to combat DeepFake, it is necessary to understand the reasons for their existence and the technologies underlying them, therefore, within the framework of the study, the following tasks were set to achieve the goal:

- To analyze the technologies of creation and application of DeepFake technology.
- To identify the advantages and threats that DeepFake technology brings to the media and society.
- To analyze the perception of fake media by modern online users.

Economic-mathematical, structural-logical research methods were used in the work, analytical processing of various documents, regulations, decrees, and laws was carried out. The use of survey methods made it possible to study and conduct a sociological study of the perception of information content by online users.

Materials and Methods

DeepFake became known to the public in 2017, when videos were posted on Reddit by users in which celebrities are compromised by false sexual situations (Oluwole et al., 2021). At the same time, it is important to note that DeepFakes are difficult to

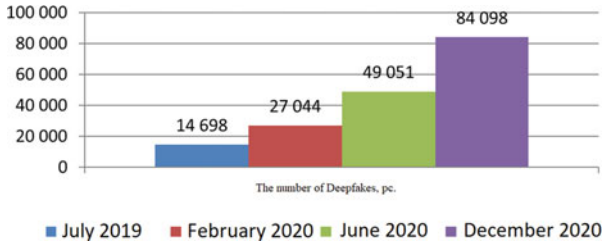

Fig. 1 The number of DeepFakes for the period 2019–2020 years

detect, since they use real frames, can have authentic sound, and are also optimized for high-speed distribution on social networks and video platforms, where conspiracies, rumors and misinformation are easily spread, since users tend to "go with the crowd".

As a result of the combination of multidirectional digital capabilities, most viewers assume that the video being viewed is authentic. At the same time, the ongoing "infopocalypse" encourages people to think that they cannot trust information if it does not come from the social networks they use. At the same time, it is interesting to note that the opinion of family members, close friends or relatives receives a lower degree of trust than social networks, bloggers, online video platforms.

Today, fake videos are common in all spheres of modern society and digital communications, this is since inexpensive equipment and affordable technologies for processing video objects are used in their creation (Vizoso et al., 2021).

In addition, technologies for creating realistic DeepFakes for disinformation purposes are becoming increasingly available, which allows users with initial technical skills to edit videos, change faces and synthesize speech. DeepFake technologies are a symbiosis of generative adversarial networks (GANS), when two artificial neural networks work together to create realistic media.

When creating DeepFake, artificial neural networks are used, which are called "generator" and "discriminator", they simulate the same set of data that are presented in the image, video data and audio data. The "generator" creates new imitated good images, and the "discriminator" works to determine how real the objects and videos of the newly created media are. GAN can view thousands of photos of a person and create a completely new portrait. Soon, generative adversarial networks will be "trained" in the presence of less information to "replace" not only the object's face, but also the entire body and voice.

Even though DeepFake usually requires a large number of images to create realistic sub-cases, researchers have already developed a method for creating fake videos by providing generative adversarial networks with only one photo, for example a selfie. The results of the analysis of the conducted studies showed that for the period 2019–2020 the number of DeepFakes has increased by 84% (Fig. 1), which indicates an exponential increase in their number.

According to the results of the study, it was revealed that, according to experts, the introduction of a legislative ban on the creation and distribution of DeepFake will only worsen the problem. As an example, confirming the validity of this statement,

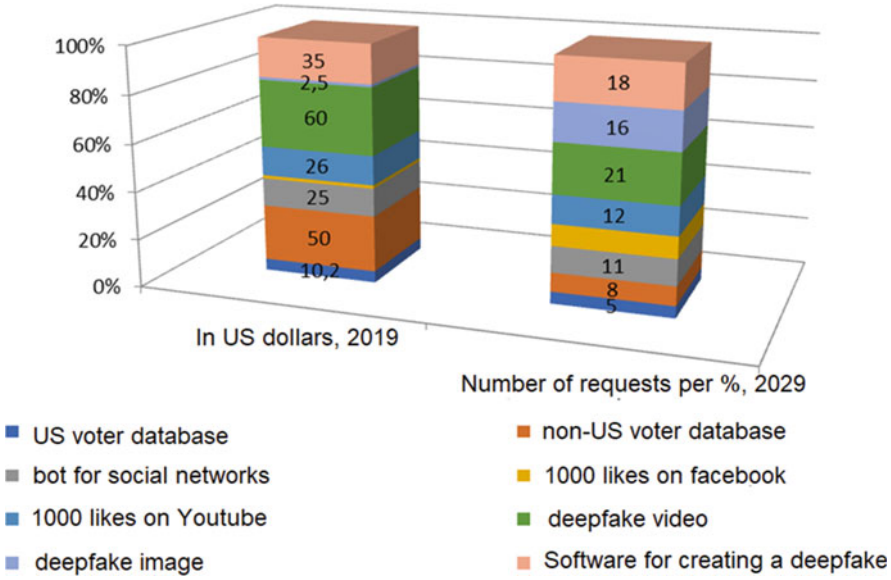

Fig. 2 The level of demand for "illegal data and information products" on the "black" market, 2019 year

experts point to China, where in 2019 fake videos were declared illegal, which allowed the Chinese government to attribute any "inconvenient photo or video" to the category of fakes. According to TrendMicro, a global leader in corporate data protection and cybersecurity solutions, in 2019 DeepFake was in the greatest demand on the "black market", as evidenced by the prices for these products (Fig. 2).

It is important to note that scientists from the SAND Lab at the University of Chicago, after analyzing statistical data, claim that the total monetary losses of global business due to DeepFake approached the $250 million mark in 2019, whereas in 2018 such losses amounted to $176 million.

Consider the positive aspects of the application of DeepFake technology, which have a positive practice of application in many industries, including cinematography, educational media and digital communications, games and entertainment, social networks, healthcare, materials science, various business areas such as fashion and e-commerce. The film industry can benefit from DeepFake technology in many ways. For example, DeepFake technology in cinematography allows you to create digital voices for actors who have lost their voice due to illness or provides updating of the film material instead of re-shooting it. Cinematographers, using DeepFake, can recreate classic movie scenes, create new films with the participation of long-dead actors, using special effects, advanced face editing in post-production and improving amateur video in professional quality.

Deepfake technology allows you to produce voice-over of different types of video objects automatically and efficiently in any language, which allows audiences from different countries to freely watch foreign films and use foreign educational media.

In 2019, as part of an information campaign about malaria with the participation of David Beckham, the issue of language barriers for the audience was resolved by placing educational ads that used visual and voice-changing technologies to obtain the "multilingualism" function. Also, an example of a positive application of DeepFake technology is overcoming language barriers at international video conferences with simultaneous translation of speech and simultaneous changes in facial and mouth movements to improve the quality of eye contact and obtain the effect of participants "speaking the same language".

The technology underlying deepfakes allows you to play multiplayer games and create virtual chat worlds with increased telepresence, creating a natural sound and types of "smart digital" assistants and digital doubles of people, which allows you to qualitatively improve human relations and interactive interaction at large distances. DeepFake technologies can have positive applications in social and medical fields. Deepfakes can help people cope with the loss of loved ones by digitally visualizing the deceased and thereby potentially help grieving people overcome the pain of loss and gradually adapt to the situation. In addition, this technology can digitally recreate a disabled limb or even help people with Alzheimer's disease.

In the context of the global development of digital technologies, researchers are studying the possibility of using GAN to detect anomalies in X-rays and their potential to create virtual chemical molecules for acceleration. Commercial global companies are interested in the potential of DeepFake technology in relation to brands, since it can significantly and qualitatively change e-commerce and advertising (Bendas, 2021).

At the present stage of digitalization of society, four main types of DeepFake producers are distinguished: communities of DeepFake lovers, political players such as foreign governments, and various activists, detractors, and criminal communities such as fraudsters, legitimate entities such as television, media companies, etc.

In general, fans of video content tend to view videos and products created by artificial intelligence as a new form of online humor and a contribution to the development of such technologies as solving an intellectual puzzle, and not to deceive or intimidate people. In this case, the audience considers entertaining, humorous, satirical, and political DeepFakes. Some of the audience is looking for more specific personal benefits, such as raising awareness about the potential of DeepFake to get a high-paying job, for example, related to music videos or TV shows. Thus, both amateurs and legitimate participants, such as TV companies, can cooperate with each other in the process of using DeepFake for various purposes.

While "meme-like" DeepFakes from amateurs can entertain Internet users, "intruders" are also involved in the process of using this technology. Various political players, including political agitators, activists, terrorists, can use DeepFake in disinformation campaigns to manipulate public opinion, to undermine confidence in a particular institution of the country, etc. Therefore, now, in addition to its positive application, DeepFake acts as a weapon of disinformation aimed at interfering in elections, inciting civil unrest, national discord, and other negative global phenomena. At the same time, DeepFake is increasingly being used by fraudsters to carry out market and exchange manipulations, as well as other serious financial

crimes. Now, there are already examples when criminals have already used fake audio recordings created by artificial intelligence to impersonate a manager by phone with a request for an urgent money transfer. At the same time, DeepFake video calls for criminal purposes (fraud, financial crimes, etc.) will be available soon, which can be made in real time. So, now, DeepFake technology can already use visual and audio impersonation of managers, for example, from videos available on TED Talk on YouTube.

Results

As part of the study, a survey was conducted of more than 300 specialists in the field of PR, marketing, advertising, and journalism. During the study, it was found out that the majority of respondents—41%—receive information through online media and information sites, social networks were the second most popular—37%. More than half of the respondents, 58%, trust information content. About 43% of respondents doubt the information presented in the videos, they are looking for additional facts and primary sources of news. Thus, it turns out that among Russian online users, more and more people are beginning to think about the reliability of information content. The videos that cause the most doubt among the online audience include videos with the following signs: the absence of an official source, a dubious or absurd title and a little-known source. The results of research are presented in Fig. 3.

In addition to the above criteria, information that promotes the imposition of a certain point of view and the presence of questionable facts, as well as pressure on negative emotions and dramatization is questioned.

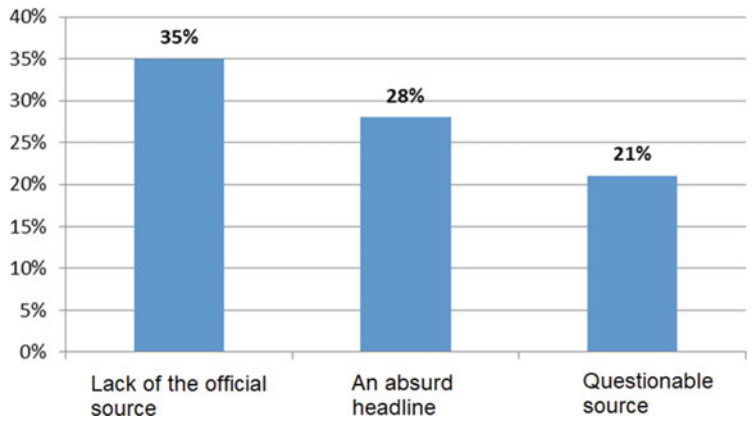

Fig. 3 The main signs of unreliability of information clips according to the online audience in 2021, according to the authors' research

Discussion

An analysis of the practice of legal methods to combat the negative effects and consequences of DeepFake indicates that now four main ones have been formed: legislative regulation, corporate policy and voluntary actions of commercial and non-profit organizations in countering negative DeepFake, digital education and training, improving digital literacy of society, technology for detecting and preventing the spread of DeepFake, authentication of content. Legislation and regulation are obvious means of preventing negative DeepFake.

Currently, DeepFake is not always considered in civil or criminal legislation, although experts in the field of world law have suggested adapting existing laws to cover libel, disinformation, identity fraud or impersonating government officials with the help of fakes. For example, the law of the US state of Virginia against revenge porn recently legislated responsibility for the distribution of "falsely created" images and videos and recognized them as an offense.

Nevertheless, the dynamic development and increasing level of complexity of artificial intelligence technologies requires new types of laws and regulatory framework with the sphere of regulation of DeepFake practice. For example, fakes raise concerns about privacy and copyright, since the visual images of people on DeepFake videos are not exact copies of any existing material, since it can rather be interpreted as new images created by artificial intelligence. Thus, regulators must navigate the complex legal landscape around freedom of speech and property laws in order to properly regulate the use of DeepFake technology.

On the other hand, a proper legal solution to the problem of spreading negative deepfakes should not be a complete ban on this technology. Today, firms working in social networks are widely using tools for using the content posted by users on their websites. One of the options for legislative regulation may be the abolition of the legal immunity of social media companies in relation to the content posted by their users, thus making not only users, but also platforms more responsible for the posted materials. At the same time, it is important to note that modern law currently has at least insufficient influence on "malicious actors", such as terrorists, who can conduct mass disinformation campaigns against other states, nationalities and nations, human values on social networks and online video platforms.

Corporate policy and voluntary actions of commercial and non-commercial organizations in countering negative DeepFake can become a more effective tool against "fakes". For example, politicians and top managers of global business giants can commit themselves not to use illegal digital communication campaigns and commercial and non-profit organizations in countering negative DeepFake tactics in election, information, market campaigns.

From the point of view of working with social media, commercial companies need to ensure compliance with ethical standards and refuse to place at the top of the content those video materials that cause global controversy in the global market (Nazaykin, 2020). Now, few social networks have a policy in relation to countering negative DeepFakes. But, in our opinion, it is social networks that should cooperate

to prevent the use of commercial and non-profit organizations in countering negative DeepFake as a weapon for disinformation, and actively introduce transparent, common political ideas to block and remove DeepFake.

Currently, many companies do not remove controversial content, but simply lower its rating to make it harder to find, making it less noticeable in users' news feeds. The increase in the amount of hostility, fake news and disinformation polluting digital platforms has forced some companies to take more active actions, such as suspending user accounts and investing in rapid detection technologies. For example, Facebook cuts off any content that is recognized as false or misleading by third-party fact-checking organizations, does not allow commercial and non-profit organizations to place ads in opposition to negative DeepFake. Instagram's algorithms will not recommend people to view content that is marked as "false" by fact-checking organizations. Among news media companies, the Wall Street Journal and Reuters have formed corporate teams to help and train their reporters to identify fake content.

Conclusion

To create an eco-friendly media space based on improving the interaction of current digital media technologies and the needs of modern society, it is necessary to do the following points:

1. To analyze technologies for creating and applying fake news that threatens people's mental and physical health.
2. To identify the main types of DeepFake and groups of their producers, analyze their goals, objectives, and applications for further neutralization.
3. To search for effective and useful ways of positive application of DeepFake technologies in modern media and society.
4. To analyze online users of digital media behavior, improve the information security of digital content.
5. To implement methods of combating DeepFake technologies in digital media, such as: legislative regulation, development of technologies for DeepFake detection, content authentication and digital methods to prevent the spread of DeepFake.

References

Afanasyeva, T., Iljina, I., & Fetisova, G. (2019). Publicity of non-profit projects as an element of sustainable territorial development. *Conference: MTSDT 2019 – Modern Tools for Sustainable Development of Territories*. Special Topic: Project Management in the Regions of Russia. https://doi.org/10.15405/epsbs.2019.12.05.70.

Bendas, D. (2021). Modern media resources in the communication system of an accordionist. *National Academy of Managerial Staff of Culture and Arts Herald*. https://doi.org/10.32461/2226-3209.4.2021.250247.

Bondarchuk, B., & Kotliar, S. (2020). News in modern media discourse. *Bulletin of Kyiv National University of Culture and Arts Series in Audiovisual Art and Production, 3*(1), 29–37. https://doi.org/10.31866/2617-2674.3.1.2020.202652

Borchard, G. (2018). *The Press and the making of modern media*. A Narrative History of the American Press (pp. 224–240). https://doi.org/10.4324/9781315658667-15.

Kambur, E. (2021). *Emotional intelligence or artificial intelligence: Emotional artificial intelligence*. https://doi.org/10.17932/IAU.FCPE.2015.010/fcpe_v07i2004.

Mirzajonov, A. (2021). The importance of discourse and media text in modern media. *ACADEMICIA an International Multidisciplinary Research Journal, 11*(2), 24–26. https://doi.org/10.5958/2249-7137.2021.00317.7

Nazaykin, A. (2020). Online resources in modern media relations. *Theoretical and Practical Issues of Journalism, 9*(4), 660–672. https://doi.org/10.17150/2308-6203.2020.9(4).660-672

Oluwole, D., Collence, C., & Tinyiko, V. (2021). *Flattening the curve of fake news in the epoch of infodemic: An epistemic challenge*. Deep Fakes, Fake News, and Misinformation in Online Teaching and Learning Technologies (pp. 143–161). https://doi.org/10.4018/978-1-7998-6474-5.ch007.

Vizoso, A., Vaz-alvarez, M., & Garcia, X. (2021). Fighting Deepfakes: Media and internet giants’ converging and diverging strategies against hi-tech misinformation. *Media and Communication, 9*(1), 291–300. https://doi.org/10.17645/mac.v9i1.3494

Yumasheva, I., & Grigorieva, O. (2021). Approaches to the classification of media theories. *UNESCO World Heritage in the hands of young people* (pp 258–262).