Buletinul Ştiinţific al Universităţii Politehnica Timişoara
Seria Limbi moderne

Scientific Bulletin of the Politehnica University of Timişoara
Transactions on Modern Languages

Vol. 23, Issue 1 / 2024

Deepfake and Media Ethics

Sorin SUCIU*

Abstract: The advancement of artificial intelligence in a manner similar to the evolution of the living world through adaptation to the social environment and survival of the fittest algorithms has led to the emergence of species of programs that challenge the way media products are made and used. The very classical structure of mediated communication, and with it the perception of the public, is now altered by the interposition of generative software. An ecology of the products made with the help of these programs and an ethics that establish milestones on the road to technological advancement in public communication are evidently necessary.

Keywords: deepfake, media ethics, artificial intelligence, truth, manipulation, distortion

1. Introduction

One of the shortest characterizations of the deepfake phenomenon was provided by Jeremy Kahn, a Bloomberg tech reporter. To him, deepfake is “kind of fake news on steroids" (excerpt from Bloomberg, 2021). The destructive potential of this phenomenon is immense in the conditions of a post-truth world in which social networks play a decisive role in the formation of public opinion. Deepfakes are evolutionary products of social networks because the same computer technology that made them possible, contributed to the creation of software that essentially distorts the audio-visual truth and propagates it with the help of social media.

As in the case of other terms that designate cutting-edge technological products, the term *deepfake* has no equivalent in Romanian. It is an expression resulted from the combination of the English terms *deep learning* and *fake*. The literal translation is "profoundly false" but it does not take into account the media specificity of this phenomenon, nor does it reflect the meaning of the combination of terms from the English language. The term is usually translated by means of the syntagmatic

*Lecturer, PhD., Department of Communication and Foreign Languages, Faculty of Communication Sciences, Politehnica University of Timișoara, Romania, sorin.suciu@upt.ro

equivalent "falsified content with the help of artificial intelligence". The phenomenon refers to the use of algorithms to replace the identity of one person with that of another person in a video, audio or photo material, thus obtaining falsified but highly credible content intended to mislead and manipulate the audience . For this purpose powerful computer tools such as machine learning, neural networks, autoencoders or generative adversarial networks are used.

Some of the most famous deepfakes are President Zelenski ordering the Ukrainian army to surrender to Russians, President Obama insulting a Republican figure and Putin declaring martial law and ordering military mobilisation. Fortunately, unlike Cambridge Analytica, another use of technology for nefarious purposes, they have not yet created problems and have not had political, electoral effects. However, they may constitute computer crimes when they are used with the purpose of misleading, manipulating, blackmailing or slandering, that is, when they affect the fair functioning of society.

In essence, the deepfake phenomenon involves an identity theft. It is not only about the visual identity, (i.e. the appearance which is not limited to the face, but can include gestures, walking, tics of that person) but also about its audio identity, namely the voice of the person with all its modulations. These constitute the authenticity features of a person. In any type of mediated electronic communication, the image accompanied by someone's recognizable voice is the guarantee that we are really dealing with that person. Due to the large amount of data that needs to be processed to create this illusion, deepfake is limited, at least for now, to recordings. A live deepfaked appearance, such as an online video chat, is not yet possible.

The society we live in is not only mediatic, but also a mediated one. This means that a large part of our experiences are not immediate, but mediated by the computer. The term is considered here in a generic sense, mobile, portable devices having a greater weight than the classic office computer. Many of the interactions between people take place in the virtual environment. Companies like *Meta* propose completely virtual environments where participants can form working groups and offer an interaction in the virtual space (which is in fact a non-space), with the help of virtual glasses, pretending to provide a "total" experience and thus claiming to take the place of the face to face interactions. Almost any human activity, be it professional, educational, leisure, entertainment, etc. is facilitated by or carried out with the help of the Internet. For the human brain, this means that two ontologically different worlds are connected, creating a continuity. The stimuli used by our brain to identify real-world objects are of the same type as those used in the virtual world. Of course, as in any evolutionary environment, predators (hackers, scammers, etc.) have appeared in the virtual world, and in response defense strategies have been developed, accompanied by increased caution on the part of users. However, just like in the real world, users are not a homogenous mass, but extremely motley. The criterion of their delusion is none other than their level of education, the knowledge they have and which makes them susceptible or not to falling into the net of cybercriminals.

2. Deepfakes and today's media

Deepfake maintains a substantial link with consumerism and media sensationalism on the one hand, and with the cultivated ignorance of the audience on the other. Sensationalism itself is an unethical media practice. A sensational news story is not one that primarily interests the public, but one that captures attention through its unusual, out of the ordinary, shocking character. The media that cultivates this type of addressing (the term information is inappropriate in this context) aims to achieve not an act of relevant communication, but one intended to induce emotional states. After viewing, the viewer is not edified, but moved, stunned, bewildered, or even shocked. Thus, things are inversed. Instead of the world becoming more intelligible, more coherent, rendered through the lens of sensationalism, it becomes more entropic, irrational and inexplicable. A long line of accidents, catastrophes, conflicts (mundane mockeries or international threats), betrayals, crises, tensions and worries. On the other side of the emotional register, the glamorous modern life of celebrities offers the public pleasures by proxy. Sensational content streams have the characteristics of addictive substances. The horror, the rage, the ecstasy, the delight need to be confirmed and relived, the hormonal glands need to be continuously tickled to produce the state of permanent stimulation and to keep the audience connected. This unethical type of mass communication achieves its goals (manipulative and mercantile, political or economic) by shaping, distorting, constructing not only contents, but also ways of presenting them. Because the sensationalist rhetoric cannot anesthetize its victims without using an entire toolkit of mobilizing means borrowed from totalitarian political propaganda: declamations, sententious phrases, serious tones, allusive music full of meanings that accentuates the images, these being commented in such a way that leaves no room for individual reflection.

What is the unethical nature of media sensationalism? In the first place, it deviates from the ideal of honest journalism. The purpose of authentic journalism is the objective information of the reader or viewer, bringing him or her up to date with the latest developments in the main domains of reality: social, economic, political, cultural, scientific, etc. The values under which the journalistic approach is placed are those of objectivity, impartiality, honesty, correctness, professionalism, courage, fidelity. Deviations from these cardinal landmarks have dramatic consequences for the public as these media sources create a complex of ignorance and misinformation. Not infrequently they are politically regimented, which explains their bias and the side angle from which they look at social reality and account for it. The contents are oriented from the perspective of this angle, which determines the construction of a parallel world that has in common with the real one the events, but not their meaning. What results from this is a dangerous product: a space of ideologically engaged communication, one with a precarious ontology in which the mystery of conspiratorial fabrications, the simplifications of an otherwise complex world, the engagement of frustrations become the marks of a media discourse of post-truth.

Deepfakes are used to obtain sensational content with the aim of discrediting people, misinforming and spreading untruths. Videos are created in which public figures say

things that they did not actually say, perform actions that they did not actually do, display behaviors that in reality do not belong to them. Deepfake videos and images present people in poses that shock, arouse amazement but, at least for a certain audience, not disbelief. As Abhishek Gupta said, “fake content is tailored toward people having fast interactions with very little time spent on judging whether something is authentic or not” (apud. Pratap, 2022). The presumption of authenticity is given by the very nature of the medium through which this information (images, video, voice) reaches us. People invest photos and videos with truth value as long as the stylistic conventions used convey that they not represent artistic films or manipulated photos (Suciu, 2019: 32). A photo is a copy of reality and if there are no signs that it has been altered by staging or post-processing, we consider it "real". The same is the case with a video footage.

3. Deepfake and Artificial Intelligence

Generative artificial intelligence refers to algorithms that have been developed to create new content such as audio, video, images, text, code, etc. Here are just a few examples of the applications made with the help of artificial intelligence that can have a significant impact on the perception of the truth transmitted through media channels:

1. Apps like Lensa can be used, even against the stated intent of their developers, to fabricate nude images of real people;
2. Voice generators trained with machine learning software, such as Play.ht by podcast.ai, were used to produce fake conversations with celebrities. Thus, Joe Rogan published a podcast in which he interviewed Steve Jobs, although they never met (Harrison, 2022);
3. DeepBrain AI offers those interested the opportunity to chat with deceased relatives whose avatar has been created and pre-trained with the help of the recorded images and voices of those who are gone;
4. Starting from a set of real images, commercial photos can be made with imaginary scenography that fits the context desired by the client, thus eliminating the costs of travel, make-up, costuming, location rentals, etc. The initial photos are uploaded to an artificial intelligence program that can place the model in any visual context;
5. Interactive AI chatbots can simulate discussions with historical figures such as Einstein, Ghandi, Charles Manson, Princess Diana or Hitler. These fictional creations can be dangerous for those users lacking critical thinking who may believe and propagate false claims (Andrew, 2023);
6. Musician David Guetta used generative artificial intelligence to craft lyrics in the style of another artist, Eminem. Then, he used a software that simulates Eminem's voice in which he uploaded the respective lyrics. The result, believable to the point of confusion, was included as a chorus in one of his songs (Keller, 2023).

7. The American media company BuzzFeed has announced that it will use artificial intelligence chatbots to generate informative content, something that could massively affect the jobs of the editors of the online media company and could create a future that some have called "dystopian".

Of course, we do not intend in a Luddite manner to dispute the considerable benefits that humanity can derive from the use of artificial intelligence. Software based on artificial intelligence can be used to efficiently and quickly realize product design solutions, to optimize industrial processes, to make medical diagnoses or to pilot autonomous vehicles. Their potential is huge and the fields of application practically unlimited. In the media they can be used to filter fake news and information, detect deepfakes and illegal, offensive or dangerous content. They can also provide a personalized experience to users by identifying their preferences and giving them suggestions based on them. With the help of tags (metadata tagging), it is possible to identify and classify the huge amounts of content constantly created and posted online or one can generate subtitles for video materials in real time.

However, as Dana Rao points out, there is enough cause for concern: "as AI can be far more powerful than human editing, very soon we will no longer be able to distinguish fact from fiction, reality from artificial reality" (apud, Pratap, 2022). At the end of the day, artificial intelligence, generative algorithms or editing software are just man-made tools. Like any tool (and AI is one of the most powerful) it can be used for noble purposes or, on the contrary, in destructive ways.

4. Media ethics and deepfake products

Among professional journalists there is unanimous agreement on the toxicity of the deepfake phenomenon which just adds to the list of manipulative influences in the media that distort information and unsettle the public. The value it serves is not truth, but confusion, ambiguity and conflict. Some of the social media platforms have taken measures in terms of combating fake-news and deepfake phenomena. Occurrences of such posts may be flagged, reported and ultimately deleted. But more is needed than that. Measures are needed to combat the spread of deepfakes by establishing a clear line of demarcation between what is permissible and what is unacceptable and reprehensible. Thus, deepfake cannot be considered just an innocent form of entertainment, parody or art. There is a considerable difference between public figure humor, political parody, artistic collage and deepfake fabrications. If in the case of the former the humorous or artistic intention is obvious, deepfakes are instead created and spread in order to substitute reality. It is a mistake to consider them innocent constructions, a figment of a rich imagination, evidence of the advance of digital computational technique.

An ethics of the virtual media that combats the deepfake phenomenon should contain the following provisions:

1. It is mandatory to respect copyright and intellectual property. Content belonging to other people may not be used without their explicit permission.

They cannot be used to create deepfakes. The right to privacy and image of people must be protected.

2. Any digital manipulation derived from an original material must state that the product is a fake and must not be interpreted as representing reality.
3. Deepfake products should never be presented as news or documentary material. Presenting them as what they are, namely digital manipulations starting from original materials, is mandatory and must leave no room for interpretation. Thus, it is required that they be labeled as fakes.
4. The creators of deepfakes must assume both the authorship of their products and the responsibility for the negative impact they have on the public in general or on certain individuals in particular.
5. Deepfakes cannot be used in any form in particular fields such as politics, advertising, sports, documentary film, etc., where they can function in a persuasive manner.

5. Conclusion

In establishing truth and separating it from falsehood, there were several stages and criteria. The criteria were religious (appeal to an allegedly infallible authority), ideological (partisan, nationalist or racist theories), pragmatic (utility establishes truth), scientific (objectivity) and now technology, still a daughter of science, offers us the possibility to falsify the very face of truth, namely the image. Just as one can document a history of truth, one can also imagine a history of falsehood and misrepresentation. The deepfake is just the latest stage in this evolution of deception, one mediated by digital technology and artificial intelligence. One can ask whether deepfake still has some virtues or is acceptable in certain areas. Our opinion is that it hasn't. The artistic or parodic uses of deepfake products can only be those in which the limits of truth are clearly fixed, and the meaning of the work of art or the parody is acquired through an unequivocal reference to them. This means that the deepfake can be included among the pathologies of communication, among fake news and post-truth.

References

1. Bloomberg. 2021. "It's Getting Harder to Spot a Deep Fake Video". Video, August, 2023. https://www.youtube.com/watch?v=gLoI9hAX9dw, Date accessed 5 September 2023.
2. Harrison, Maggie. 2022. "Deepfaked Podcast has Joe Rogan Interview undead Steve Jobs. Futurism, October, 13, 2022. https://futurism.com/the-byte/deepfaked-podcast-joe-rogan-interview-steve-jobs, Date accessed 17 November 2022.
3. Keller, Erin. 2023. "David Guetta deepfakes Eminem's voice in new song: 'Future of music is in AI'." New York Post, February, 2023, https://nypost.com/2023/02/14/david-guetta-deepfakes-eminems-voice-in-new-song-future-of-music-is-in-ai/, Date accessed 23 Mars 2023.

4. Paul, Andrew. 2023. " 'Historical' AI chatbots aren't just inaccurate – they are dangerous". Popular Science, January, 26, 2023. https://www.popsci.com/technology/historical-figures-app-chatgpt-ethics/, Date accessed 5 September 2023.
5. Pratap, Aayushi. 2022. "Deepfake Epidemic is Looming – And Adobe is Preparing For The Worst". Forbes, June 29, 2022. https://www.forbes.com/sites/aayushipratap/2022/06/29/deepfake-epidemic-is-looming-and-adobe-is-preparing-for-the-worst/?sh=3f3152b75b81, Date accessed 24 August 2022.
6. Suciu, Sorin. 2019. "Communicating Through Photographic Images". *Professional Communication and Translation Studies*, Vol. 12, 2019: 29-33.