THE RISK OF ARTIFICIAL INTELLIGENCE FOR DEMOCRACY AND THE EU'S FIRST EFFORTS TO REGULATE IT

Jaroslav Denemark[*]

Abstract: *The disinformation techniques are evolving alongside with the evolution of artificial intelligence. The creating and dissemination of disinformation is more effective especially due to perpetrators using deepfake as convincing illusion of reality, bots as artificial means of making disinformation relevant and psychographic microtargeting as powerful tool for tailoring perfect disinformation for each individual. European Union is aware of the dangers imposed by artificial intelligence, but the question is whether current or proposed regulations have the potential to tackle the issue effectively.*

Keywords: *Artificial Intelligence, Artificial Intelligence Act, Digital Services Act, Disinformation, Deepfake, Bots, Psychographic Microtargeting*

INTRODUCTION

Even though European Union has fully recognized disinformation as a problem for democracy in 2015,[1] the need to tackle this issue was never so blatant as it is now. It is, apparently, due to the massive outburst of social media usage and inevitable (but phased) shift from the "offline" reality to the "online" virtual reality.

The problems that can be caused by disinformation were more than ever obvious during pandemic of COVID-19.[2] Needless to say that the activities of third countries such as China and particularly Russia aimed to undermine "democratic debate, exacerbate social polarization and improve their own image."[3] Hence, the debate regarding COVID-19 and vaccinations was substantially influenced via disinformation produced and spread by third parties to strengthen their position by provoking and stimulating distrust of the citizens towards the national states and European Union as a whole.[4]

The dangers of disinformation for democracy are even enhanced by the use of artificial intelligence, as is clearly recognized by the European Union in the European Democracy Action Plan, where is artificial intelligence primarily associated with the psychographic microtargeting, powerful tool for psychological profiling of the disinformation recipients.[5]

[*] JUDr. Jaroslav Denemark, Department of European Law, Faculty of Law, Charles University in Prague, Prague, Czech Republic. ORCID: 0000-0002-8304-1312. "This output was produced in the framework of the Specific University Research Project (***Specifického vysokoškolského výzkumu) 2023-260 619***)".

[1] Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Action Plan against Disinformation, 5 December 2018, JOIN(2018) 36 final, pp. 2–3.

[2] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions On the European democracy action plan, COM(2020) 790 final, 3. 12. 2020, pp. 18–19.

[3] Ibidem, pp. 19–20.

[4] Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Tackling COVID-19 disinformation – Getting the facts rights, 16 June 2020, JOIN(2020) 8 final

[5] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – "On the European democracy action plan", 3 December 2020, COM(2020) 790 final, pp. 2.

Artificial intelligence makes it simpler for anyone not only to create disinformation, but also to spread it, where the cost efficiency relates not only to effort and time but also to "price" of disinformation,[6] which ultimately makes it dangerous tool for democracy if used with *male fidei*.

This paper aims to analyze the use of artificial intelligence for the purpose of disseminating and creating disinformation. Firstly, the analysis will focus on main techniques based on the artificial intelligence that are used especially in the online space of social media such as bots, psychographic microtargeting and deepfake. After the thorough analysis of these mechanisms the article will focus on answering the question whether the currently prepared[7] proposal for the "Artificial Intelligence Act"[8] regulates analyzed mechanisms and if so, whether the proposed regulation could be sufficient enough to maintain democracy free from negative impact of artificial intelligence-based disinformation.

I. PERILS OF ARTIFICIAL INTELLIGENCE

Influence on the citizens can be surprisingly straightforward trough political disinformation spread during the election campaigns. Most recently, European Commission was worried how Russia will interfere during the parliamentary election campaign in Slovakia and whether this interference could shift the attitude of the public towards populistic parties or parties that explicitly incline to Russian ideology.[9] For this reason, European Commission Vice President Věra Jourová, who is the lead person of EU digital affairs, European democracy and transparency, had arranged meeting ahead of the Slovak elections with leading (social) media companies Alphabet (Google), TikTok and Meta.[10]

According to Věra Jourová, the purpose of the meeting was to assure Slovak officials and European Commission representatives that the online platforms providers will follow their voluntary commitments contained in the Code of Practice on Disinformation.[11,12] In other words, the providers assured Slovak officials and European Commission that they try to mitigate impact of disinformation on the Slovak elections and to mitigate their influence

[6] CASERO-RIPOLLÉS, A., TUNÓN, J., BOUZA-GARCÍA, L. "The European approach to online disinformation: geopolitical and regulatory dissonance". Humanities & Social Sciences Communications, pp. 3. In: *nature.com* [online]. [2024-01-31]. Available at: <https://doi.org/10.1057/s41599-023-02179-8>.

[7] But still not in the final wording.

[8] Proposal for a Regulation of the European parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 21 April 2021, COM(2021) 206 final, 2021/0106 (COD).

[9] GOUJARD, C. "Specter of Russian interference hangs over Slovak election, EU warns". In: *politico.eu* [online]. 26. 9. 2023 [2024-01-31]. Available at: <https://www.politico.eu/article/european-union-slovakia-elections-russian-interference-disinformation/>.

[10] SCOTT, M. "TikTok and Meta warned over Slovakia elections lies". In: *politico.eu* [online]. 25. 9. 2023 [2024-01-31]. Available at: <https://www.politico.eu/article/alphabet-tiktok-meta-slovakia-election-digital-services-act/>.

[11] Interview with Věra Jourová In: *European Commission* [online]. 16. 6. 2022 [2024-01-31]. Available at: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>; <https://www.youtube.com/watch?v=UZRX33wChhU&t=145s&ab_channel=Pr%C3%A1vnick%C3%A1fakultaUK>.

[12] The 2022 Strengthened Code of Practice on Disinformation [online]. [2024-01-31]. Available at: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>.

over the citizens. Even though the extent to which the Slovak elections were influenced by disinformation is yet to be determined, it is now clear that the winning party SMER was highly supported by such disinformation coming from Russia as the party's attitude towards Ukraine favors the Russian narrative.[13]

The victory of Rober Fico, leader of SMER, may represent weakening of unity in the European Union in supportive measures aimed to aid Ukraine during the Russian invasion and may lead in overall increase of anti-EU voices.

Concerns of deliberate support of SMER in campaign by Russian-originated disinformation are becoming even more real after first steps done by newly appointed Slovak government of populist party SMER and social democrats HLAS. On 3 November 2023 Juraj Gedra – newly appointed head of governmental office – has issued video on his Facebook page[14], where he proclaims that it is necessary to get rid of the "political activists" in the governmental office, Ministry of Defense, Ministry of Interior Affairs and Ministry of Foreign Affairs in order to prevent those "activists" from dictating the Slovak "the only allowed truth". Those "political activists" are employees of the Slovak republic who work on agenda focused on disclosure and fight against Russian propaganda and disinformation. Dozens of employees are now prevented from doing their work, e.g. Ministry of Foreign Affairs no longer draws attention of the public to Russian propaganda, head of Centre for Combating Hybrid Threats Daniel Milo was made redundant as well as administrator of the Slovak police Facebook page David Púchovský, who was *inter alia* highlighting and debunking disinformation in online space.[15]

II. DEEP FAKE: NEW UNREAL REALITY

However, the Slovak elections were memorable not only for historically recorded highest number of disinformation spread via social media during election campaign, but also due to the use of particular disinformation techniques based on artificial intelligence.[16]

Firstly, there was a very popular video shared on social media, especially Facebook, where voices of Zuzana Čaputová – president of the Slovak Republic, and Michal Šimečka – leader of liberal and progressive party Progresívne Slovensko claimed their support for extremist right-wing nationalist party Republika while abandoning their liberal values.[17,18]

[13] GOUJARD, C. "Specter of Russian interference hangs over Slovak election, EU warns". In: *politico.eu* [online]. 26. 9. 2023 [2024-01-31]. Available at: <https://www.politico.eu/article/european-union-slovakia-elections-russian-interference-disinformation/>.

[14] GEDRA, J. "Za Úřad vlády SR potvrdzujem, ze politickí aktivisti už nebudú diktovať Slovákom povinne správný názor". In: *Juraj Gedra, facebook.com* [online]. 3. 11. 2023 [2024-01-31]. Available at: <https://www.facebook.com/story.php/?id=100080174506997&story_fbid=336909962324818&refsrc=deprecated&paipv=0&eav=AfYreEwEvARsCEjK40mwafpd3oBu-V3lhLpShl244hW5ni_qWhkbpmMhoY9hkzC9o6k&_rdr#_=_>.

[15] MAREK, J. "Ficova vláda otřásá Slovenskem. Odchází odborníci na ruské vlivové operace a šíření dezinformací". In: *iRozhlas.cz* [online]. 22. 11. 2023 [2024-01-31]. Available at: <https://www.irozhlas.cz/zpravy-svet/slovensko-robert-fico-odbornici-na-boj-s-ruskymi-dezinformacemi-ruske-vlivy_2311221204_elev>.

[16] ZMUŠKOVÁ, B. "Progressive Slovakia becomes target of AI misinformation, tops polls". In *EURACTIV* [online]. 28. 9. 2023 [2024-01-31]. Available at: <https://www.euractiv.com/section/politics/news/progressive-slovakia-becomes-target-of-ai-misinformation-tops-polls/>.

[17] Ibid.

[18] Saying that he "used to believe in 70 genders and pregnant man."

But Šimečka, as a leader of liberal opposition to nationalist parties, was target of deepfakes more than once. Another deepfake video showed Šimečka talking with most famous Slovak investigative journalist Monika Tód; Šimečka (or deepfake version of Šimečka's voice) proclaims that he wants to buy votes from Roma minority group in order to win elections. Furthermore, on the voice recording Šimečka and Tód are discussing a plan how to influence the election results through illegal manipulation and by corrupting other journalists.[19] Although Šimečka and Tód alongside with many factcheckers publicly rejected anything said on the alleged recording and artificial intelligence systems stated that the voice recording is deepfake, it was already subject to avalanche effect reaching hundreds of thousands of people.

Even more alarming is fact that the deepfake showed Šimečka and Tód joking and laughing about child pornography.[20] Another very popular deepfake video was the voice recording, where Šimečka plans to increase beer prices after his party wins the elections.[21] Even though it is not yet certain how did those particular deep fakes influenced the election result, it is clear that for the first time in history of EU was artificial intelligence used in such a magnitude and with direct intent to influence the future composition of the political representation.[22]

Generally speaking, deepfake is virtual audio and/or graphic (video, image) content created or manipulated to appear as "authentic or truthful"[23] or simpler put, deepfakes depict "individuals saying and doing things they never said or did."[24] The danger of deepfake consist mainly of incapacity of individuals to detect them and, consequently, distinguish between the reality and falsehood.

The key problem is that people generally tend to rely on their intuition and associative processes rather than on the cognitive processes.[25] This theory was confirmed by study focused on the perception of visual media (in the case the study of pictures). The study showed that when presented with pictures that were in some way manipulated, people tend to believe that they even witnessed the faked news demonstrated by the doctored picture.[26] This phenomenon is naturally only amplified in case of video or even voice recording, where people almost subconsciously believe that what they see or hear is reality.

Even though deepfakes that show high level of quality in resolution, e.g. clear picture, undisturbed voices, are still relatively unique, so called "shallowfake" or "cheapfake" are

[19] KUNDRA, O. "Jak snad zmanipulovat volby". In: *RESPEKT* [online]. 21. 10. 2023 [2024-01-31]. Available at: <https://www.respekt.cz/tydenik/2023/43/jak-snadno-zmanipulovat-volby>.

[20] ZUIDIJK, D. "Deepfakes in Slovakia Preview Howw AI Will Change the Face of Elections". In: *Bloomberg* [online]. 4. 10. 2023 [2024-01-31]. Available at: <https://www.bloomberg.com/news/newsletters/2023-10-04/deepfakes-in-slovakia-preview-how-ai-will-change-the-face-of-elections>.

[21] Ibidem.

[22] VERMA, P., OREMUS, W. "AI voice clones mimic politicians and celebrities, reshaping reality." In: *The Washington Post* [online]. 15. 10. 2023 [2024-01-31]. Available at: <https://www.washingtonpost.com/technology/2023/10/13/ai-voice-cloning-deepfakes/>.

[23] Proposal for a Regulation of the European parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 21 April 2021, COM(2021) 206 final, 2021/0106 (COD), Article 52 Sec. 3.

[24] APPEL, M., PRIETZEL, F. The detection of political deepfakes. *Journal of Computer-Mediated Communication*. Vol. 27, No 4, [2024-01-31]. Available at: <https://doi-org.ezproxy.is.cuni.cz/10.1093/jcmc/zmac008>.

[25] Ibidem.

[26] Ibidem.

flooding the online space. Even in the case of alleged recordings of Martin Šimečka, the voices were not as clear as "official" recordings, and the poor quality was to demonstrate that it was recorded on mobile phone or from far. Shallowfakes (cheapfakes) are even more hard to ban as they might not be fully fabricated but can only slow the speech or show that the speech is in some way slurred.[27] This was the case of for example manipulated video of Nancy Pelosi, U.S. House Speaker, which showed her as intoxicated or sick.[28] Even though this particular video was quickly debunked by US officials, it was viewed more than 2.5 million times on Facebook in just a few days after release and was shared even by many politicians.[29]

The case of shallowfake appeared also during Czech presidential elections, where the doctored video of then candidate and currently Czech president Petr Pavel showed him saying that the only possible way of how to help Ukraine is to declare war to Russia. This statement was however the exact opposite of what he actually said.[30]

The biggest issues with deepfakes is the fact that they aim to "undermine the credibility of legitimate information"[31] making the fake information alternative to the reality or the reality as something that is fake.[32]

III. BOTS AS BOTH QUANTITATIVE AND QUALITATIVE TECHNIQUE TO SPREAD DISINFORMATION

Disinformation is most credible, when it is are spreading massively or is interacted with, either in a way of approval or disapproval. If the disinformation is not shared or is without reactions (comments, "like" button etc.), it means that it has no public reach and therefore is ineffective. To artificially create the impression that disinformation is interacted with or shared from the beginning (of its existence in public space), fake accounts on social media, or bots, are used. In other words, bots are mainly used for amplification of disinformation.[33]

Bots can be described as "algorithmically driven computer programs designed to carry out specific tasks online such as analyzing and scraping data",[34] which means that bots are usually used by the online services providers such as Google bot. However, these programs can be also "created to automatically post content, increase followers' numbers,

[27] STASI, M. L., PARCU, P. L. "Chapter 21: Disinformation and misinformation: the EU response". In: Pier Luigi Parcu – Elda Brogi (eds.). *Research Handbook on EU Media Law and Policy.* Cheltenham: Edward Elgar Publishing, 2021.

[28] "Doctored Nancy Pelosi video highlights threat of 'deepfake' tech". In: *CBS NEWS* [online]. [2024-01-31]. Available at: <https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/>.

[29] APPEL, M., PRIETZEL, F. *The detection of political deepfakes.*

[30] BIDRMANOVÁ, M., *"Sítěmi se valí lživé video o generálu Pavlovi. Je za ním proruský účet."* In: *Seznam Zprávy* [online]. 22. 1. 2023 [2024-01-31]. Available at: <seznamzpravy.cz/clanek/domaci-sitemi-se-vali-lzive-video-o-generalu-pavlovi-je-za-nim-prorusky-ucet-224034>.

[31] BONTRIDDER N, POULLET Y. *The role of artificial intelligence in disinformation.* In: *Data & Policy* [online]. 2021 [2024-01-31]. Available at: <https://www.cambridge.org/core/journals/data-and-policy/article/role-of-artificial-intelligence-in-disinformation/7C4BF6CA35184F149143DE968FC4C3B6>.

[32] Ibidem.

[33] STASI, M. L., PARCU, P. L. *"Chapter 21: Disinformation and misinformation: the EU response".*

[34] Ibidem, pp. 410.

support political campaigns and spread misinformation or disinformation."[35] The power of those bots is that they are *perceived* as real persons, so when bot comments or shares content, it seems that the content is commented or shared by living, autonomously thinking human. Therefore, bots are both quantitative, as there can potentially be infinite number of them, and qualitative, as they are on the first sight unrecognizable from real person, tool for dissemination of disinformation.

Those particular bots described above are labelled as "social bots", as they are programed to interact with users on social media.[36] Sub-class of social bots are political bots programed to support politicians, campaigns or specific political agenda or idea.[37] Social bots are then categorized into two basic groups as "chat bots" and "spam bots", where those bots based on key word searches react to post of social media users by either creating conversation ("chat bots") or by sharing specific content ("spam bots").[38]

That bots are threat to democracy is clearly demonstrated by data showing that far-right parties (usually supported by Russia) use bots to promote their topics twice as much as is average across all parties.[39] To analyze the activity of bots during 2019 European elections, British journal The Independent developed ana algorithm that found that 12 % of all tweets on (then) Twitter promoting and supporting far-right parties originated from automated fake accounts – bots. Only in 2 days, bot accounts created and tweets supporting Matteo Salvini, leader of popular far-right party in Italy, more than 400 tweets.[40] Furthermore, origins of most of the bots linked to far-right politicians can be tracked to Russia.[41] In the same European elections, analysis said, more than 6 700 bots reached approximately 241 million EU citizens, which is around half of all the EU citizens.[42]

The threat of bots is also well recognized by the EU institutions. European Union Agency for Cyber Security in its annual Threat Landscape Report raise concerns over chat bots and their role in disinformation campaigns and even directly links this threat to upcoming 2024 EU elections.[43] This warning of European Union Agency for Cyber Security was backed by its annual report only show that bots should not be underestimated as they might present, when used with *malae fidei*, immanent threat for democracy.[44]

[35] Ibidem.

[36] BRKAN, M. Artificial Intelligence and Democracy: The Impact of Disinformation, Social Bots and Political Targeting. *Delphi – Interdisciplinary Review of Emerging Technologies*. 2019, Vol. 2, No. 2, p. 67.

[37] Ibidem.

[38] Ibidem.

[39] BEVENSEE, E., ROSS., R., A., NARDIN, S. We built an algorithm to track bots during the European elections – what we found should scare you. In: *The Independent* [online]. 22. 5. 2019 [2024-01-31]. Available at: <https://www.independent.co.uk/voices/european-elections-parliament-bots-social-media-matteo-salvini-far-right-a8924831.html>.

[40] Ibidem.

[41] BOFFEY, D. 241m Europeans' may have received Russian-linked disinformation. In: *The Guardian* [online]. 8. 5. 2019 [2024-01-31]. Available at: <https://www.theguardian.com/world/2019/may/08/241m-europeans-may-have-received-russian-linked-disinformation?fbclid=IwAR1ovG3LvfVc0ecFAeGW10DoBlhk3VnY4elK0DtQTtfTL3fHD_lQrV3dauk>.

[42] Ibidem.

[43] ROUSSI, A. European election at risk from AI, says EU's cyber agency. In: *politico.eu* [online]. 19. 10. 2023 [2024-01-31]. Available at: <https://www.politico.eu/article/european-union-election-risk-artificial-intelligence-interference-cybersecurity-agency-enisa/>.

[44] ENISA Threat Landscape 2023. In: *ENISA* [online]. 19. 10. 2023 [2024-01-31]. Available at: <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>.

IV. WHEN THE PERPETRATOR KNOWS EVERYTHING – PSYCHOGRAPHIC MICROTARGETING

Psychographic microtargeting is a product of so-called surveillance capitalism.[45] Online services providers (especially social media and search tools providers) usually work on free-to-use basis. Users do not have to pay for using Instagram, Facebook (Meta Inc.) or Google search (Alphabet Inc.), at least not with official currency. Those online services providers profit from offering advertising products and services of third parties on the platforms they administrate. This form of marketing is most effective, when the content of each user's interface is tailored to his needs, desires, age, social group, gender etc.

Online services providers have access to almost infinite amount of data about each user. Google Analytics, tool for tracking movement on e-shops, is in form of cookies almost on every website you visit. Facebook plugins are almost the same case. Data retrieved by the online services providers are then used to customize what users see and interact with. These practices can pose imminent danger to personal data processing and were (not once) scrutinized by European Court of Justice.[46] The ultimate goal is to create advertisement for each user based on his personality.

Psychographic microtargeting is done by algorithms, usually artificial intelligence driven computer programs whose role is to gather and process information and create output on which basis the content is personalized.[47] This algorithm can be also used for political campaigns, where data such as church attendance, type of car purchases, education and membership of Facebook groups can be relevant to make a picture of "possible" voter.[48] On the basis of the profile, person sees tailored political advertisements that tackle his personality. Needless to say, that social media enables this form of targeting more than any other platform.[49]

It is important to note that there is a difference between demographic and psychographic targeting (profiling). Demographic profiling is well known technique in sociology and used by politicians throughout the history. Demographic profiling enables to segment the voter based on "age, education, employment, country of residence and in some cases ethnicity.[50] For example, based on the demographic profiling, former Czech president candidate Danuše Nerudová targeted younger electorate which was later targeted by President Petr Pavel as well.[51] Furthermore, based on the demographic profiling, analyst can

[45] DAWSON, J. Microtargeting as Information Warfare. *The Cyber Defense Review.* 2021, Vol. 6., No. 1, p. 67.

[46] Judgment of the European Court of Justice, *C-252/21 – Meta Platforms Inc., Meta Platforms Ireland Ltd., Facebook Deutschland GmbH v. Bundeskartellamt.* In: *European Court of Justice* [online]. 4. 7. 2023 [2024-01-31]. Available at: <ECLI:EU:C:2023:537>.

[47] DAWSON, J. *Microtargeting as Information Warfare*, p. 67.

[48] Ibidem, p. 68.

[49] NAPOLI, P., M., GRAF, F. Chapter 4: Social media platforms as public trustees: an approach to the disinformation problem. In: Taina Pihlajarinne – Anette Alén-Savikko (eds.). *Artificial Intelligence and the Media.* Cheltenham: Edward Elgar Publishing, 2021, p. 94.

[50] KERTYSOVA, K. Artificial Intelligence and Disinformation: How AI Challenges the Way Disinformation Is Produced, Disseminated, and Can Be Countered. *Security and Human Rights.* 2018, Vol. 29, No. 1-4, p. 64.

[51] ŠPANĚLOVÁ, K., HEJL HROMKOVÁ, D. Pavel se snaží přetáhnout Nerudové mladé voliče. Pil s nim zelenou a hrál fotbálek. In: *Aktuálně.cz* [online]. 6. 1. 2023 [2024-01-31]. Available at: <https://zpravy.aktualne.cz/domaci/pavel-kampan-mladi/r~40b82aca8dae11edbc030cc47ab5f122/>.

identify that Czech populist and former presidential candidate Andrej Babiš' electorate consists of elder people, people with only basic education and people living in poor regions, on the other hand Petr Pavel's electorate consist of educated people, who have better living conditions.[52] This profiling shows, on what groups will politicians most likely focus. However, psychographic profiling enables to focus not only on groups, but on each individual based on whether the individual is for example extrovert or introvert, what are the popular purchase items, what are his opinions on different political agenda etc.[53]

Best example of use of psychographic microtargeting was possibly provided by Cambridge Analytica[54] during 2016 presidential elections and during Brexit campaign. Cambridge Analytica, officially for research purposes, created psychographic personality profiles of 100 million registered voters in the USA and use it to target specific campaign segments of Donald Trump.[55] During Brexit campaign, Cambridge Analytica was leading consulter of Leave.EU, where Cambridge Analytica used algorithmic tools to profile potential voter for "leave".[56] During the Brexit "leave" campaign, Cambridge Analytica primarily focused on "Left Behind" group of people who felt "increasingly left behind by society and globalization", were "unhappy with the economy and the NHS, but immigration is most important issue", were "suspicious of the establishment including politicians banks and corporations" and were "worried about their economic security, deterioration public order and the future generally".[57] Data about those potential voters were illegally sourced by Cambridge Analytica from UKIP database, Facebook, insurance companies and other personal data processors.[58] Based on those data, Leave.EU could target disinformation on immigration, NHS and EU on each individual based on his profile in order to convince him to vote for leaving the EU.[59]

Furthermore, social media platforms are prioritizing false content over factual content as false content means more interactions and therefore can be better monetized.[60] This "economic of attention"[61] makes the most interacted content available to even more users, as the algorithm seeks to target as many users as possible in order to make bigger profit.[62]

Now, it is important to remind the role of "social bots" – to artificially make disinformation seem to be interacted with, shared and commented. If the algorithm recognizes

[52] GRIM, J. V obcích s vysokou mírou vzdělání vyhrál Pavel. V místech s nadprůměrným počtem exekucí zase Babiš. In: *iRozhlas.cz* [online]. 14. 1. 2023 [2024-01-31]. Available at: <https://www.irozhlas.cz/volby/volby-2023-prezidentske-andrej-babis-petr-pavel_2301141519_jgr>.

[53] KERTYSOVA, K. *Artificial Intelligence and Disinformation: How AI Challenges the Way Disinformation Is Produced, Disseminated, and Can Be Countered.* p. 64.

[54] Cambridge Analytica was consulting firm focusing on profiling of persons in order to foster campaign of its clientele.

[55] KERTYSOVA, K. *Artificial Intelligence and Disinformation: How AI Challenges the Way Disinformation Is Produced, Disseminated, and Can Be Countered.* p. 65.

[56] BAKIR, V. Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting. *Frontiers in Communication.* 2020, p. 10.

[57] Ibidem.

[58] Ibidem.

[59] Ibidem, pp.11–12.

[60] KERTYSOVA, K. *Artificial Intelligence and Disinformation: How AI Challenges the Way Disinformation Is Produced, Disseminated, and Can Be Countered.* p. 64.

[61] BONTRIDDER N, POULLET Y. *The role of artificial intelligence in disinformation.* pp. 4.

[62] Ibidem, p. 5.

the content as highly interacted with even by social bots, it automatically spread this content to more users.[63]

V. SOLUTION ON THE HORIZON (?)

As demonstrate hereinabove, use of artificial intelligence for disinformation campaigns is immanent danger for democracy. Furthermore, artificial intelligence is used more then ever before due to the rapid ascent of this type of technology. It is, in my opinion, therefore crucial to regulate these types of artificial intelligence, so that algorithms are transparent and use of artificial intelligence is clearly labelled so that the general public has no doubts about the origin of video, image, voice recording or social media account.

This approach is in general proclaimed by the European Commission in its Communication "Tackling online disinformation: a European approach".[64] First of all, Commission points out that in order to tackle disinformation, transparency should be significantly improved, i.e. origin and the way the information is "produced, sponsored, disseminated and targeted"[65] in order to reveal "possible attempts to manipulate opinion".[66]

Other steps to fight against disinformation in general are according to Commission diversity of information (quality journalism, free media etc.), to foster credibility of information (by trusted flaggers or generally by fact checking to provide indications of trustworthiness) and to fashion inclusive solutions (education, raising awareness etc.).[67] However, this particular Commission's communication is not focused on the issues of artificial intelligence *per se*, except briefly mentioning role of bots to "artificially amplify the spread of disinformation".[68]

Even though the abovementioned Commission's communication was not binding, several largest online services providers such as Meta Inc., Alphabet Inc. and later TikTok (while Twitter was originally proactive as well, it withdrew from this activity later on) agreed on creating and signing "Code of Practice on Disinformation", which was reviewed and strengthened in 2022.[69] In the Code of Practice on Disinformation, signatories represent various commitment they ought to fulfill in order to effectively combat against disinformation. Commitment 15 of the Code is dedicated solely to "Transparency obligations for AI systems". In this Section of the Code, signatories voluntarily refer to yet just a proposal for Artificial Intelligence Act especially committing to take into consideration transparency obligations.[70] Furthermore, they commit (beyond obligations contained in the artificial intelligence act proposal) to implement op-

[63] Ibidem.

[64] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Tackling online disinformation: a European Approach*, 26 April, 2018, COM(2018) 236 final.

[65] Ibidem, p. 6.

[66] Ibidem.

[67] Ibidem.

[68] Ibidem, p. 5.

[69] The Strengthened Code of Practice on Disinformation 2022. In: *European Commission* [online]. 16. 6. 2022 [2024-01-31]. Available at: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>.

[70] Ibidem, p. 17.

erating mechanisms to detect and prevent further dissemination of disinformation (explicitly in the form of deepfakes).

But the Code of Practice on Disinformation is still only voluntary documentation signed only by interested parties without possibility to enforce the contained obligations. This, however, might change when the proposed Artificial Intelligence Act becomes effective.

The proposal is now before Council of the European Union in first reading, and it is possible that the text will be changed.[71] However, some of the Articles of the proposal can be already scrutinized in order to preliminary estimate whether the Artificial Intelligence Act might regulate three biggest threats of artificial intelligence use in disinformation warfare.

The centerpiece of the proposal is transparency of AI systems. In Article I let. (c) of the proposal, where subject matter is laid down, it says that: "This Regulation lays down: (…) harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorization systems, an AI systems used to generate or manipulate image, audio or video content."[72]

This transparency rules effect use of artificial intelligence in a way that the addressee is aware of it. This is one of the most important rules in order to avoid artificial intelligence techniques to lead disinformation discourse.

More specifically, proposal make it mandatory for "providers"[73] to "ensure that Ai systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an Ai system, unless this is obvious from the circumstances and the context of use (…)."[74] This section of the proposal clearly aims to AI systems used as mobile operators, various costumer support systems on e-shops and other Ai systems, where the purpose or intend of the system is not malicious. It is, however, unclear whether bots (in a meaning hereinabove – especially social bots) would be regulated by this provision as well. Meta or (formerly) Twitter are not providers of artificial intelligence programs (algorithms) that operate the fake accounts and therefore are not responsible for labeling such fake accounts as bots. It would be naïve to think that the perpetrators (trolls – troll farms) would label their bots as artificial intelligence in order to be transparent. These obligations shall be laid upon the online services providers as well. It should be mandatory to implement such mechanisms that allow the (particularly) social media providers to recognize and effectively eradicate fake accounts – bots.

Moreover, using bots in a way described hereinabove is not even prohibited by the AI Act proposal. Among prohibited artificial intelligence practices are, however: "the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond person's consciousness in order to materially distort a person's behaviour

[71] For example Věra Jourová told in personal interview noted that the category of so called „high-risk AI systems" might be expanded for AI systems used for political campaigns.

[72] Article 1, let. (c) of the proposal.

[73] Where provider means pursuant to Article 3, Sec. (2) a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developer with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge.

[74] Article 52, Sec. 1 of the proposal.

in a manner that causes or is likely to cause that person or another person physical or psychological harm".[75] In a way, using bots is subliminal technique, where the human user perceives the fake account as a real person and therefore subliminally treats the bot's actions as such. However, it cannot be said that the use of bots for the purpose of disseminating (and creating) disinformation causes physical or psychological harm which makes this prohibition unusable in this context.

On the other hand, proposal explicitly refers to deep fake, where it is mandatory for "users"[76] of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ("deep fake"), shall disclose that the content has been artificially generated or manipulated(…).[77]

This however means that not the service provider, but the user (meaning the one who generates or uploads the content) shall label it as deep fake. It might help with dissemination of disinformation by official political parties, who are under public scrutiny, it does not, however, ensure that third parties that are nearly untraceable will obey such obligations.

This means that Artificial Intelligence Act proposal does not make it mandatory directly for social media providers to label content as created by artificial intelligence or to actively search for and possibly delete fake accounts operated by bots. As for the psychographic microtargeting, proposal introduce minimum to no rules related to that matter regarding disinformation.

It is, however, possible that the social media provider will be obliged to monitor compliance with the rules set out by the Artificial Intelligence Act proposal pursuant to Digital Services Act.[78] Under the Digital Services Act, illegal content means "any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of the law."[79] This would mean that any content, with regard to disinformation and artificial intelligence in most cases deepfakes, breaching rules under Artificial Intelligence Act would be considered as illegal pursuant to Digital Services Act.

Pursuant to Digital Services Act, the social media providers (as referred in the Digital Services Act online platforms) have obligations relating to monitoring illegal content, implementing reporting obligations, internal complaint-handling system and to transparency reporting obligations.[80]

Digital Services Act also regulates psychographic microtargeting, but only with relation to the providers of intermediary services (including online platform) and their algorithms

[75] Article 5, let. (a) of the proposal.

[76] Where "user" means pursuant to Article 3, Sec. (4) of the proposal any natural or legal person, public authority, agency or other body using an Ai system under its authority, except where the AI system is used in the course of a personal non-professional activity.

[77] Article 52, Sec 3 of the proposal.

[78] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Service Act, L 277/7.

[79] Ibidem, Article 3, let. (h).

[80] Even though Digital Services Act explicitly excludes general monitoring obligations.

used for personalization of content for the users. Those obligations consist mainly of transparency and reporting rules.[81] This, however, does not tackle the problem of using psychographic microtargeting by malicious third parties in order to target specific types of disinformation as effectively as possible. Thus, only EU regulation impacting on psychographic microtargeting is GDPR,[82] which regulates only the processing such as collecting and use of personal data, but not the actual issue of using psychographic microtargeting for spreading disinformation. Furthermore, Digital Services Act does not provide any regulation regarding bots.

CONCLUSION

Even though that artificial intelligence poses many opportunities for technological improvement, it also presents threat for the current state of democracy. Deep fakes, bots and psychographic microtargeting are techniques that thrive on each day technologically sophisticated artificial intelligence.

Although it should be leading interest of the European Union to regulate the artificial intelligence technology that helps to create and disseminate disinformation more effectively, the current state of *acquis* is insufficient. Artificial Intelligence Act proposal explicitly regulates only deepfakes and their use by "users", and further regulates specific AI systems that exploit vulnerabilities or work on subliminal manipulation. However, prerogatives imposed by artificial intelligence act on situation in which use of those specific types of artificial intelligence are prohibited are too narrow and might not impact the regulation used for creation and dissemination of disinformation at all.

Digital Services Act might help to combat disinformation by imposing certain obligations for the social media providers regarding illegal content, it does not, however, provide complex solution, when the issues of bots and use of psychographic microtargeting by malicious third parties is not regulated at all.

The European Union should therefore introduce complex regulation focused solely on disinformation and responsibility of online platforms for combating them. Until then we can only hope that voluntary Code of Practice on Disinformation will be followed.

[81] Article 14, Sec. 1 of the Digital Services Act.

[82] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regards to the pricessing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).