ESSAYS II/2022

OBSAH SEKCE

REASONS FOR BIAS IN AUTOMATED DECISIONS AND POTENTIAL REMEDIES[1]

ANNA BLECHOVÁ[2]

1. INTRODUCTION

The use of automated decision-making systems in judicial practice is becoming more frequent in recent times. For example, Mexico is using a tool called EXPERTIUS, "a *decision-support system that advises Mexican judges and clerks upon the determination of whether the plaintiff is or not eligible for*

[1] Esej byla zpracována v semestru podzim 2021 v rámci předmětu MVV1368K Privacy and Personal Data na téma Automatic decision making. / The essay was written in the autumn 2021 semester for the course MVV1368K Privacy and Personal Data on the topic of Automatic decision making.

[2] Anna Blechová je studentkou magisterského studijního programu Právo a právní věda na Právnické fakultě Masarykovy univerzity, kontakt: 458594@mail.muni.cz

granting him/her a pension",[3] Estonia developed and piloted an AI-based (automated) system to hear and decide on specific claims disputes[4] and some US judges can use a risk-assessment tool called COMPAS ('Correctional Offender Management Profiling for Alternative Sanctions') which is also based on an automated process and helps to infer which of the convicted defendants is most susceptible to recidivism to decide about bail or sentence.[5]

Since automated decision-making in judicial practice is a relevant issue I decided for the purposes of this essay to focus on its pitfalls. Specifically, I will focus my attention on *bias* as one of the main threats of automated decision-making systems based on AI or machine learning. To narrow down the given topic, this essay will answer the main research question; (i) *Are automated decision-making systems in judicial practice biased?* If the answer to the main research question will be affirmative two sub-questions will follow: *(ia)Are there any potential remedies to this issue?*, and *(ib)Is it appropriate to use the remedies?*.

The essay will be structured in the following way: After the Introduction (I) the notion of (non)bias in automated decision-making tools will be presented (IIa). Afterwards, the reason for bias will be explained (IIb), and furthermore, some potential remedies to this issue will be submitted (IIc). In the following part, the main attention will be paid to the question, whether it is appropriate to use any remedies against bias in automated decision-making tools (IId). Ultimately, based on the previously mentioned, the Conclusion (III) will recapitulate and summarize the answers to the research questions.

[3] CARNEIROA, Davide et al. Online Dispute Resolution: an Artificial Intelligence Perspective. Artificial Intelligence Review. [online]. vol. 2014, no. 41. [cit. 20. 11. 2021] s. 227–228.

[4] NIILER, Eric. Can AI Be a Fair Judge in Court? Estonia Thinks So. *Wired* [online]. 2019. [cit. 20.11.2021]. Available at: https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/

[5] ZALNIERIUTE, Monika. Technology and the Courts: Artificial Intelligence and Judicial Impartiality. *SSRN Electronic Journal*. [online]. 2021. [cit. 20. 11. 2021]. DOI: 10.2139/ssrn.3867901

2. (NON)BIASED DECISION-MAKING TOOLS

2.1 IS IT THERE OR IT IS NOT THERE, THAT IS THE QUESTION

The first crucial question which should be answered is if automated decision-making tools are biased or not. To answer this question it is necessary to primarily define what bias is. According to the Merriam Webster dictionary, bias is defined as *"a tendency to believe that some people, ideas, etc., are better than others that usually results in treating some people unfairly"*.[6]

It follows from the above mentioned that one of the conceptual features of bias is the *tendency to believe*. But can machine learning or AI, a piece of technology, *believe* in anything? The ability to believe is connected to thinking. One of the options to evaluate if the automated decision-making process can think is via the Turing test, also known as the" imitation game".[7] This test is simple. It is based on three variables – variables A, B and C, one of them is human (A or B), the second one is a computer (A or B), and the last one is a tester (C, human). The computer aims to "convince" the human tester that it is also a human, not a computer. If the machine is successful and outsmarts the human being, it is concluded that it can think.[8] There were several attempts[9] to pass the test, but until today nobody successfully completed it.[10] Inasmuch as there was no successful attempt to pass the Turing test, machines cannot think; thus, they cannot actually "be" biased.

[6] Definition of BIAS. In: *Merriam-Webster dictionary.* [online]. c 2021. [cit. 20.11.2021]. Available at: https://www.merriam-webster.com/dictionary/bias

[7] TURING, Alan.—Computing Machinery and Intelligence. *Mind.* 1950, vol. LIX, no. 236. DOI: 10.1093/mind/LIX.236.433

[8] SHAH, Raivat. Can Machines Think? In: *Medium* [online]. 17. 11. 2019. [cit. 20.11.2021]. Available at: https://towardsdatascience.com/can-machines-think-307e16e3fd2c

[9] AAMOTH, Dough. Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test. In: *Time* [online]. 9. 6. 2014. [cit. 20.11.2021]. Available at: https://time.com/2847900/eugene-goostman-turing-test/

[10] PANOVA, Evgeniya. *Which AI has come closest to passing the Turing test? - Dataconomy* [online]. 2021. [cit. 20.11.2021]. Available at: https://dataconomy.com/2021/03/which-ai-closest-passing-turing-test/

Nevertheless, there are examples of automated decision-making tools within the judicial procedure that seem to be biased. One of the examples is the risk assessment tool for criminal cases from the US, which are based on hard data from questionaries, criminal records etc., which appears biased to the detriment of the black people.[11] This phenomenon was described in detail in the report by ProPublica.[12]

2.2 BE BIASED VS. APPEAR BIASED

How is it possible that automated decision-making tools appear biased although they are not capable of thinking? The answer to this question will be divided into three parts each referring to a problematic aspect.

The first problematic element is the human being itself. The reason is, that humans are the creators of the systems. Moreover, since people, as the fundamental element of the process of creating automated decision-making tools are biased, the system and especially the data fed to the system may be biased. Besides, society actually expects something from machines in which it fails itself. Bias, which relates to prejudice, is a feature of human beings, even of judges for example. This has been proven by the research by Danzinger, who found out that judges after having a meal are more moderate in their decision-making than judges who are hungry.[13]

Another facet that can contribute to the bias is data. According to Surden, the algorithms are "*only as good as the data that they are given to analyse".*[14] It is important to understand that data that is fed to the discussed tools are not a 1:1 reflection of the real world. Moreover, they cannot even

[11] HAO, Karen. AI is sending people to jail—and getting it wrong. In: *MIT Technology Review.* [online]. 21. 1. 2019. [cit. 20.11.2021]. Available at: https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

[12] ANGWIN, Julia et al. *Machine Bias* [online]. ProPublica, 2016. [cit. 22.11.2021]. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[13] DANZIGER, Shai, Jonathan LEVAV a Liora AVNAIM-PESSO. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*. [online]. 2011, vol. 108, no. 17. [cit. 20. 11. 2021]. DOI: 10.1073/pnas.1018033108

[14] SURDEN, Harry. Machine Learning and Law. Washington Law Review 89 [online]. 2014, no. 87. [cit. 20. 11. 2021]. p. 106.

be. Thus, one of the problems of the dataset is its scope and quantity. Lehr is adding to this point that, the data scientist needs to be sure, that they collected *"enough data"* because running the machine learning or AI systems on small data sets is pointless.[15] Further issues with the data set are the up-to-datedness and inflexibility. The development and learning of an AI or machine learning system are complicated and long-term projects. In relation to this, developers work with a set of data that is stable and inevitably from the past.[16] This is causing a lack of reaction to the development of society and new trends. To conclude, if the system is based on limited, outdated and stable data, it cannot be accurate and reflect reality.

Ultimately, the last problematic aspect is the *"insufficient complexity"* of the systems. As was already mentioned, neither AI nor machine learning-based systems can think and, furthermore, they cannot think *"out of the box".* For example, even though the system will be based on accurate, flexible, unlimited data an unexpected variable in the computation could cause that a specific case will not *"fit in the box".*[17] This issue is based on the general approach to training of AI systems, which Haeven considers as flawed.[18]

In summary, even automated decision-making systems within the judicial procedure could appear biased. This is caused by the fact that they are not complex enough and they are created and fed by the inaccurate data produced by biased humans. Thus, the answer to the main research question is affirmative.

[15] LEHR, David a Paul OHM. Playing with the Data: What Legal Scholars Should Learn About Machine Learning. *U.C. Davis Law Revie*. 2017, vol. 52, no. 2, p. 677–678.

[16] SURDEN, Harry. *Machine Learning and Law*, p. 105.

[17] LEHR, David a Paul OHM. *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, p. 711.

[18] HEAVEN, Will Douglas. The way we train AI is fundamentally flawed. In: *MIT Technology Review* [online]. 2020. [cit. 22.11.2021]. Available at: https://www.technologyreview.-com/2020/11/18/1012234/training-machine-learning-broken-real-world-heath-nlp-computer-vision/

2.3 POTENTIAL REMEDIES TO BIAS IN ADM SYSTEMS

The potential remedies to the bias will be for the purposes of this text divided into two parts. The first part will focus on the data, the second one on suitable policies.

As was already mentioned, the algorithm is as good or biased as are its learning data. According to this, one of the possible solutions for mitigating bias is to be precise with creating the dataset. Moreover, it is not only the data itself but also the team which is selecting them. Thus, it is crucial to avoid the lack of diversity in programming teams because it can lead to the under-representation of a particular group or specific physical attributes.[19] Another issue is the data timeliness. The solution for this problem seems simple – use the up-to-date data. However, this may be difficult in practice. In my opinion, even if the data were outdated, the careful selection and mitigation of the problematic assets could overcome it. Nevertheless, the question is what outdated truly means. If outdated means that the data are from the previous decade, it would be more problematic than if the data are a month or two old. To this point it is quite important to add, that the data collection takes some time.

To mitigate bias in algorithmic decision-making is also possible by the use of precise legal or policy frameworks. A ban of algorithmic discrimination may be easier to enforce and easier to convey to the victim of the infringement of such a specific right. Moreover, legislative actions can provide guardrails that are applicable when automated decision-making tools caused harm.[20] An example of such legislative action could be the European Union *"Ethics Guidelines for Trustworthy AI"*.[21]

[19] BARTON, Genie, Nicol TURNER a Paul RESNIK. *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms* [online]. 2019. [cit. 22.11.2021]. Available at: https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

[20] Ibid.

[21] EUROPEAN COMMISSION. *Ethics Guidelines fot Trustworthy AI* [online]. 2019. [cit. 20. 11. 2021]. Available at: https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

In closing, there are potential remedies to bias of automated decision-making systems. Examples of such mitigation tools are meticulously created datasets and establishing legal and policy frameworks.

2.4 DO WE WANT TO DE-BIAS?

According to Celiskan, the problem with bias is, that bias and non-ambiguity is by default connected with natural languages.[22] Since the AI or machine learning system will be trained on a dataset based on natural language, which is probable in the area of law, the system will be by its very nature biased. In other words the dataset for prediction tools contains information from previous judgements, it is connected to written expressions of law, and is also based on information from questionaries which are all written in natural language and the input data are biased. Thus, to de-bias, the automated decision-making tools should not be based on natural language. Even though this could be technically possible, the question is, if the "translation" from natural language to the binary language will help. Since when we do so, we can lose some nuances in translation.

Another interesting point in this matter was raised by Završnik. He claimed in *"Algorithmic justice: Algorithms and big data in criminal justice settings"* that even our constitutions and codes *"have all been adopted through a democratic legislative process that distilled the prevailing societal interests, values, and so on of the given society."*[23] In other words, in the process of creating the rules the humans already imprinted their biases in them and thus, there is no doubt that constitutions and codes are also biased. Moreover, if the de-biased code would be implemented, the decision of correctness of the data would not be made public by the politicians

[22] CALISKAN, Aylin, Joanna J. BRYSON a Arvind NARAYANAN. Semantics derived automatically from language corpora contain human-like biases. *Science*. [online]. American Association for the Advancement of Science, 2017, vol. 356, no. 6334. [cit. 20. 11. 2021]. p. 185.

[23] ZAVRŠNIK, Aleš. Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*. [online]. SAGE Publications, 2021, vol. 18, no. 5. [cit. 20. 11. 2021]. p. 633.

(and society) but by the top-level IT expert behind closed doors. Thus, the democratic element will evaporate. It is really the desired effect?[24]

Oppositely, it could be claimed that technology could be used to fight bias instead of entrenching it. For example, algorithms are able to eliminate systematic bias, which could result in environments that encourage disadvantaged groups to succeed. The techniques to accomplish this goal are for example "turning off" the source of bias (for example age) or calibrating different cut-off scores.[25]

In conclusion, the answer to the last research sub question (*(ib)Is it appropriate to use the remedies?)* is from my point of view unclear. This is mainly because we can easily find arguments for both sides. Nevertheless, in my opinion, technology is more like a mirror to our society. According to that, the usage of the de-biasing tools could be an interesting approach to "be better", but it is not something that should be the ultimate argument for the damnation of the automated decision-making systems. Thus, the answer to the question posed is, that it is appropriate to use the remedies, but it is necessary to be cautious with such tools.

3. CONCLUSION

Automated decision-making systems in judicial practice are nowadays extensively used in jurisdictions all over the world. It is thus understandable, that society wants this tool to be almost flawless. Unfortunately, it is not and one of the possible problems of decision-making systems is algorithmic bias.

To sum up the findings of the essay, even though automation decision-making tools cannot think, they can appear biased. This is primarily caused by the data on which they are based. Even though the presence of bias is undeniable, there are at least two ways how to mitigate it. One of the possibilities is the remedy via an appropriately selected dataset, the other possi-

[24] Ibid.

[25] BAER, Tobias. *How Algorithms Can Fight Bias Instead of Entrench It - By Tobias Baer* [online]. 2020. [cit. 22.11.2021]. Available at: https://behavioralscientist.org/how-algorithms-can-fight-bias-instead-of-entrench-it/

bility is via legal and policy frameworks. According to the dilemma of the desirability of de-biasing the systems, it is necessary to conclude that de-biasing tools could be good tools for the improvement of the systems, but they should not be the ultimate argument for the damnation of the automated decision-making systems.

4. BIBLIOGRAPHY

[1] BAER, Tobias. *How Algorithms Can Fight Bias Instead of Entrench It - By Tobias Baer* [online]. 2020. [cit. 22.11.2021]. Available at: https://behavioralscientist.org/how-algorithms-can-fight-bias-instead-of-entrench-it/

[2] BARTON, Genie, Nicol TURNER and Paul RESNIK. *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms* [online]. 2019. [cit. 22.11.2021]. Available at: https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

[3] CALISKAN, Aylin, Joanna J. BRYSON a Arvind NARAYANAN. Semantics derived automatically from language corpora contain human-like biases. *Science*. [online]. American Association for the Advancement of Science, 2017, vol. 356, no. 6334, p. 183–186. [cit. 20. 11. 2021]. DOI: 10.1126/science.aal4230

[4] CARNEIROA, Davide et al. Online Dispute Resolution: an Artificial Intelligence Perspective. *Artificial Intelligence Review*. [online]. vol. 2014, no. 41, p. 211–240. [cit. 20. 11. 2021]. DOI: https://doi.org/10.1007/s10462-011-9305-z

[5] DANZIGER, Shai, Jonathan LEVAV a Liora AVNAIM-PESSO. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*. [online]. National Academy of Sciences, 2011, vol. 108, no. 17, p. 6889–6892. ISSN 0027-8424, 1091-6490. [cit. 20. 11. 2021]. DOI: 10.1073/pnas.1018033108

[6] AAMOTH, Dough. Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test. In: *Time* [online]. 9. 6. 2014. [cit. 20.11.2021]. Available at: https://time.com/2847900/eugene-goostman-turing-test/

[7] EUROPEAN COMMISSION. *Ethics Guidelines fot Trustworthy AI* [online]. 2019. [cit. 20. 11. 2021]. Available at: https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

[8] HEAVEN, Will Douglas. The way we train AI is fundamentally flawed. In: *MIT Technology Review* [online]. 2020. [cit. 22.11.2021]. Available at: https://www.technologyreview.com/2020/11/18/1012234/training-machine-learning-broken-real-world-heath-nlp-computer-vision/

[9] ANGWIN, Julia et al. *Machine Bias* [online]. ProPublica, 2016. [cit. 22.11.2021]. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[10] HAO, Karen. AI is sending people to jail—and getting it wrong. In: *MIT Technology Review* [online]. 21. 1. 2019. [cit. 20.11.2021]. Available at: https://www.technologyreview.-com/2019/01/21/137783/algorithms-criminal-justice-ai/

[11] LEHR, David a Paul OHM. Playing with the Data: What Legal Scholars Should Learn About Machine Learning. *U.C. Davis Law Revie*. 2017, vol. 52, no. 2, p. 653–718.

[12] NIILER, Eric. Can AI Be a Fair Judge in Court? Estonia Thinks So. *Wired* [online]. 2019. [cit. 20.11.2021]. ISSN 1059-1028. Available at: https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/

[13] PANOVA, Evgeniya. *Which AI has come closest to passing the Turing test? - Dataconomy* [online]. 2021. [cit. 20.11.2021]. Available at: https://dataconomy.com/2021/03/which-ai-closest-passing-turing-test/

[14] SHAH, Raivat. Can Machines Think? In: *Medium* [online]. 17. 11. 2019. [cit. 20.11.2021]. Available at: https://towardsdatascience.com/can-machines-think-307e16e3fd2c

[15] SURDEN, Harry. Machine Learning and Law. *Washington Law Review 89* [online]. 2014, no. 87.[cit. 20. 11. 2021] Available at: https://digitalcommons.law.uw.edu/cgi/viewcontent.cgi?article=4799&context=wlr

[16] TURING, Alan. Computing Machinery and Intelligence. *Mind.* [online]. 1950, vol. LIX, no. 236, p. 433–460. [cit. 20. 11. 2021]. ISSN 0026-4423. DOI: 10.1093/mind/LIX.236.433

[17] ZALNIERIUTE, Monika. Technology and the Courts: Artificial Intelligence and Judicial Impartiality. *SSRN Electronic Journal*. [online]. 2021. [cit. 20. 11. 2021]. ISSN 1556-5068. DOI: 10.2139/ssrn.3867901

[18] ZAVRŠNIK, Aleš. Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*. [online]. SAGE Publications, 2021, vol. 18, no. 5, p. 623–642. [cit. 20. 11. 2021]. ISSN 1477-3708. DOI: 10.1177/1477370819876762

[19] Definition of Bias. *In:* Merriam-Webster Dictionary [online] c2021 [cit. 21. 11. 2021]. Available at: https://www.merriam-webster.com/dictionary/bias

THE POTENTIAL OF SMART CONTRACTS BEYOND THE CONTEXT OF DECENTRALIZED FINANCE[1]

MARTIN ERLEBACH[2]

1. INTRODUCTION

There seems to be certain amount of hype surrounding term "smart contract" recently, not just in mainstream publications but also in academic papers spanning many scientific branches and fields of research. This in turn most probably made term "smart contract" into kind of buzzword. Proponents of smart contracts promise fantastical things but mainly disruption of legal professions, "cutting out middleman" and revolutionizing contract law all at once.

I do not believe smart contracts are able to fulfil many of promises they set out to accomplish. In this paper, will first and foremost try and define what smart contract is. Subsequently, will elaborate on connection between blockchain technology and smart contracts. will also try and describe why smart contracts will not revolutionize contract law in their current state by describing, at least briefly, which broad legal and technical hurdles would be necessary to overcome to actually deliver on what they are so often connected with.

[1] Esej byla zpracována v semestru podzim 2021 v rámci předmětu MVV57917K Regulating Disruptive Technologies na téma Blockchain. / The essay was written in the autumn 2021 semester for the course MVV57917K Regulating Disruptive Technologies on the topic of Blockchain.

[2] Martin Erlebach je studentem magisterského studijního programu Právo a právní věda na Právnické fakultě Masarykovy univerzity, Kontakt: 480066@mail.muni,cz

2. DEFINING SMART CONTRACT

Defining relatively new thing is always hard. Most of time, and smart contracts are no exception, academics hurry to develop their own definition regarding subject of study if it is something new. In case of smart contracts, definitions found in scientific literature can sometimes be simple such as "an agreement whose execution is automated" which is "effected through computer running code that has translated legal prose into an executable program".[3] On the other hand, there are much more complex definitions such as aspiring legal definition of smart contract from Arizona which states: *"Smart contract" means an event-driven program, with state, that runs on distributed, decentralized, shared and replicated ledger and that can take custody over and instruct transfer of assets on that ledger."*[4]

In end, on most basic level, most researchers can agree that smart contract is classical "if-then" statement that runs on blockchain where "parties can enter into binding commercial relationship, either entirely or partially memorialized using code, and use software to manage contractual performance."[5] Smart contracts will be understood as such within this paper. addition of running contract on blockchain is an important one since without it we could be as well talking about vending machine because it basically monitors performance of contract independently as well (when enough money is inserted and an item of that or lower price is selected it dispenses it) be it with an initial human input.[6]

If we excluded critical part about blockchain smart contracts are not such new thing after all, contrary to what was said right at top of paper. first similar thought, originally called Electronic Data Interchange (or EDI

[3] RASKIN, Max. Law And Legality Of Smart Contracts. *Georgetown Law Technology Review* [online]. 2017. [cit. 13.01.2022]. p. 309.

[4] KINTER, Eric. Arizona Authorizes Smart Contracts on Blockchain | Data Privacy and Protection Blog [online]. 4.4.2017 [cit. 13.01.2022]. Available at: https://www.swlaw.com/blog/data-security/2017/04/04/arizona-authorizes-smart-contracts-on-a-blockchain/

[5] DE FILIPPI, Primavera, WRIGHT, Aaron. *Blockchain and Law: Rule of Code*. 2018. p. 46.

[6] SZABO, Nick. Idea of Smart Contracts [online]. c 1997 [cit. 13. 1. 2022]. Available at: https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html

for short) agreements, was introduced in 1970s. They were also hyped up to disrupt face of contractual law, but eventually failed to deliver on their promises and just helped cut some costs in business.[7]

3. PROMISES OF SMART CONTRACTS

In this part of paper, would like to explore small section of different promises and revolutions smart contract proponents envision for them other than just their use in decentralized finance field. view these as sort of core proposed advantages of smart contracts which are most often mentioned when talking about them.

3.1 IMMUTABILITY

first promise regarding smart contracts as they were defined above stems from fact that they are run on blockchain, which is also known as distributed ledger technology. This technology is mostly known in connection with cryptocurrencies[8] and with huge amount of simplification (since it is not object of this paper) described as database of transactions kept simultaneously by people participating in blockchain. For purposes of smart contracts, it is most often talked about as public blockchain model which means anyone can access it if they so choose.[9] Lastly, important aspect is that this network of databases or ledgers comprises of blocks which are segments of transactions connected to each other in succession. This way technology should prevent tampering or changing anything on this blockchain network since for block or transaction to be valid it must connect to longest previous chain of blocks.[10]

[7] SKLAROFF, Jeremy. *Smart Contracts and Cost of Inflexibility*. Rochester, NY: Social Science Research Network, [online] 2017. [cit. 13. 01. 2022]. p. 274.

[8] SEGAL, David. My Puzzling Entry in Crypto World. *New York Times*, [online]. 2021. [cit. 13.01.2022]. Available at: https://www.nytimes.com/2021/08/17/insider/cryptocurrency-hype-coin.html

[9] MILLER, Andrew, DELMOLINO, Kevin, KOSHBA, Ahmed and SHI, Elaine. *Step by Step Towards Creating Safe Smart Contract: Lessons and Insights from Cryptocurrency Lab*, [online] 2015. [cit. 13. 01. 2022]. Available at: http://eprint.iacr.org/2015/460

[10] NAKAMOTO, Satoshi. Bitcoin: Peer-to-Peer Electronic Cash System, [online]. 2008. [cit. 13. 01. 2022]. p. 1.

Naturally, when you deploy smart contract which specifies obligations on such network it should be by definition immutable or said bit simpler, unchangeable.[11] In this way, you can be sure that once contract is deployed it will be executed way it was coded. This has obvious benefits like protection from falsification or change of contract without notifying other party, which is fear some might have.

3.2 TRUST-LESS ENVIRONMENT

other core trait of smart contracts should also stem from idea of blockchain technology. This upside over traditional contract is that contract can be trustless, but what does that mean exactly? Well, it means that contract should be executed when certain conditions are met, always. With traditional contracts, you must rely on other party for performance of contract (e. g. transfer of money). In this way, smart contracts are "self-executing".[12]

This should, along with immutability of contract ensure that everything around contract goes smoothly. This feature of smart contracts also plays role in "cutting out middleman" part of smart contract promises. Because if everything would go smoothly and according to smart contract, parties would be forced to act according to it by underlying code, and there would be no costs associated with need to enforce contract in any way.[13] other party cannot possibly behave in different way. This not only negates need to trust other contractual party but also trust in legal system and courts since in an ideal case there isn't any need for enforcement of contract with courts and their ambiguous rulings and uncertain outcomes.

3.3 EFFICIENT

last one of these core upsides smart contracts are supposed to have is idea that they are highly efficient. This also corresponds with aspect of "cutting

[11] GUADAMUZ, Andres, MARSDEN, Chris. Blockchains and Bitcoin: Regulatory responses to cryptocurrencies. *First Monday*, [online] 2015. [cit. 13. 01. 2022]. p. 1.

[12] GUADAMUZ, Andres, *All Watched Over by Machines of Loving Grace: Critical Look at Smart Contracts*. Rochester, NY: Social Science Research Network, [online] 2019. [cit. 13.01.2022] p. 2.

[13] Ibidem.

out middleman”. In nutshell, efficiency of smart contracts lies in fact that they are, as was mentioned above, self-executing and they do not need enforcement by outside powers. To these advantages, it can also be added that smart contracts promise to cut out even lawyers that draft traditional contracts and are often portrayed as very closed group of specialists developing their own language and maybe, just maybe driving up prices of easy tasks.

To all these advantages might add that smart contracts promise to revolutionize not just traditional legal professions by replacing traditional paper contracts with code but also promise betterment of any product to customer tracking using blockchain. This use could range from mere traceability of coffee from plant to your cup but could also be used in medicine to track transplants or marihuana for medical use which both have to be strictly monitored.[14]

In conclusion, potential for smart contracts in modern world seems to be huge and this paper barely touched on all proposed uses for this technology.

4. PITFALLS OF SMART CONTRACTS

above mentioned begs question, why has technology not been yet implemented everywhere? We can certainly blame some of this on fact that blockchain technology is itself rather young,[15] so it is not yet as readily accepted by general population.

But in this paper, it will explore what might be other reasons for this low adoption rate and why think smart contract technology is not set to change what we know about contract law outside of decentralized finance.

[14] ZINOVYEVA, Elizaveta, REULE, Raphael C.G., HÄRDLE, Wolfgang K. Understanding Smart Contracts: Hype or Hope?. Rochester, NY: *Social Science Research Network*, [online]. 2021. [cit. 13.01.2022]. p. 2.

[15] As proof we can see that bitcoin whitepaper was published in 2008, see: NAKAMOTO, Satoshi. Bitcoin: Peer-to-Peer Electronic Cash System, [online]. 2008. [cit. 13.01.2022]. Available at: https://bitcoin.org/bitcoin.pdf

4.1 IMMUTABILITY AS FLAW

first of flaws is other side of same coin we presented above. contracts are immutable. This also means that it is very hard and expensive to change them if need arises, be it from simple novation agreed with other party or worse, to repair bug or something that can be exploited. process of changing smart contract already deployed on blockchain simply put consists of taking it down and starting and re-deploying amended version of code.[16] This is of course extremely inefficient and most importantly requires agreement of all parties involved in smart contract to be executed, which can be dangerous if let's say flaw was advantageous for one of parties.

4.2 TRUST BUT IN CODE

second pitfall of smart contracts is rather easily identifiable. Smart contracts are essentially machine-readable code executed on blockchain (sometimes also called DApps).[17] And code for most part must still be written by human, which obviously comes with possible bugs[18] in code which are so hard to get rid of as was explained above. This surely goes somewhat again notion that "done by machine is better than by human" which is often associated with smart contracts.[19] Another aspect of dealing with possible faults in code is expenses one must expend to deploy smart contract on blockchain and execute it. These costs can be quite unpredictable because of different miner fees associated with different blockchains. These fees often change dynamically with demand for transaction verification or for example in case of Ethereum are caused by limitation of fees possible to get

[16] ZINOVYEVA, Elizaveta, REULE, Raphael C.G., HÄRDLE, Wolfgang K. Understanding Smart Contracts: Hype or Hope?. Rochester, NY: *Social Science Research Network*, [online]. 2021. [cit. 13.01.2022]. p. 74.

[17] State of DApps. What's DApp [online]. c 2022. [cit. 13. 1. 2022]. Available at: https://www.stateofthedapps.com/whats-a-dapp

[18] OLICKEL, Hrishi. Why Smart Contracts Fail: Undiscovered bugs and what we can do about them [online]. *Medium*. 9. 9. 2019 [cit. 13. 1. 2022]. Available at: https://hrishiolickel.medium.com/why-smart-contracts-fail-undiscovered-bugs-and-what-we-can-do-about-them-119aa2843007

[19] Ibidem.

from block.[20] This can lead to state where deployment of smart contract is more expensive than it was first thought.

second big problem with code of smart contracts is difficult readability of code for non-tech savvy people and ability to hide nefarious provisions in code.[21]

In traditional contracts, we at least have some law to protect consumers from unreadable terms. [22] Nevertheless, it is uncertain if this law could apply to smart contracts.Since in this way many of disadvantaged parties could be harmed, we could even expand list of disadvantaged groups by people who cannot read code.

4.3 DISCUSSION ABOUT LEGAL IMPLICATIONS

last point of this paper should be some of many legal implications that smart contracts have in regard to established traditional contract law. These implications often connect with concerns expressed above and may offer glimpse into reasons why smart contracts probably will not change contract law field or make lawyers or traditional contracts obsolete.

first one is possible discrepancy between actual contract law and rules included in smart contract. Let us say that applicable law entitles tenant to demand rebate of 100 % when apartment he is renting through smart contract fails to have hot water for week. But adding such provision to smart contract (essentially adding applicable law to code) would be expensive as was explained above. So, landlord opts to leave such code out of contract, or coder just forgets to add it. This could easily lead to tenant being locked out of their apartment for perfectly legal behaviour in accordance with contract.[23]

[20] Ibidem

[21] Ibidem

[22] For example Directive 2011/83/EU of European Parliament and of Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of European Parliament and of Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of European Parliament and of Council, [online]. 2011. [cit. 09.06.2022]. Available at: http://data.europa.eu/eli/dir/2011/83/oj/eng

Another fault in smart contracts can be seen in fact, that we are not yet able to transform all legal prose into code.[24] Some obligations are very much vague and purposefully so. Some authors argue that code cannot express such rules.

This is connected with one more problem that plagues smart contracts. It is so-called “oracle problem” after software that feeds real-world data into blockchain called “Oracles”. It can be described as problematic way of connecting digital world with physical one.[25] In above-mentioned example with tenant, lease contract can be easily facilitated with smart lock, but how would you correctly input finishing of roof in way that is trust less as smart contracts promise? Or if contract is supposed to force parties to do something for each other and it is not transfer of money but physical service like assembling furniture for cooked meal? believe smart contracts are not equipped yet for this kind of contracting and can serve only as form of strengthening contract in form of contractual penalty executed on blockchain.

last legal implication is idea of different voluntary prerequisites for formation of contract. Some legal actions based on existing contracts may require express will of party to be enforced. It is still not clear if smart contract automated execution could be considered as such.[26]

5. CONCLUSION

In conclusion, still believe smart contracts are, at least not yet, fit to change landscape of contractual law as we know. More likely there will be minor

[23] ZINOVYEVA, Elizaveta, REULE, Raphael C.G., HÄRDLE, Wolfgang K. Understanding Smart Contracts: Hype or Hope?. Rochester, NY: *Social Science Research Network*, [online]. 2021. [cit. 13.01.2022]. p. 59.

[24] GUADAMUZ, Andres, *All Watched Over by Machines of Loving Grace: Critical Look at Smart Contracts*. Rochester, NY: Social Science Research Network, [online] 2019. [cit. 13.01.2022] p. 2.

[25] DELPHI. Oracle Problem [online]. *Medium*. 15. 7. 2017 [cit.13. 01.2022]. Available at: https://medium.com/@DelphiSystems/the-oracle-problem-856ccbdbd14f

[26] ZINOVYEVA, Elizaveta, REULE, Raphael C.G., HÄRDLE, Wolfgang K. Understanding Smart Contracts: Hype or Hope?. Rochester, NY: *Social Science Research Network*, [online]. 2021. [cit. 13.01.2022]. p. 58.

upgrade in areas with high volume of repeated, highly formal contractual relationships, like decentralized finance, but nothing more. believe such an opinion will hold while pitfalls of smart contracts defined above apply.

6. BIBLIOGRAPHY

[1] KINTER, Eric. Arizona Authorizes Smart Contracts on Blockchain | Data Privacy and Protection Blog [online]. 4.4.2017 [cit. 13.01.2022]. Available at: https://www.swlaw.com/blog/data-security/2017/04/04/arizona-authorizes-smart-contracts-on-a-blockchain/

[2] DE FILIPPI, Primavera, WRIGHT, Aaron. *Blockchain and Law: Rule of Code*. 2018.

[3] DELPHI. Oracle Problem [online]. *Medium*. 15. 7. 2017 [cit.13. 01.2022]. Available at: https://medium.com/@DelphiSystems/the-oracle-problem-856ccbdbd14f

[4] Directive 2011/83/EU of European Parliament and of Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of European Parliament and of Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of European Parliament and of Council, [online]. 2011. [cit. 09.06.2022]. Available at: http://data.europa.eu/eli/dir/2011/83/oj/eng

[5] GUADAMUZ, Andres. *All Watched Over by Machines of Loving Grace: Critical Look at Smart Contracts*. Rochester, NY: Social Science Research Network, [online] 2019, p. 1-16. [cit. 13.01.2022] Available at: https://papers.ssrn.com/abstract=3805473

[6] GUADAMUZ, Andres, MARSDEN, Chris. Blockchains and Bitcoin: Regulatory responses to cryptocurrencies. *First Monday*, [online] 2015. p. 20-32. [cit. 13.01.2022]. Available at: https://firstmonday.org/ojs/index.php/fm/article/view/6198

[7] MILLER, Andrew, DELMOLINO, Kevin, KOSHBA, Ahmed and SHI, Elaine. *Step by Step Towards Creating Safe Smart Contract: Lessons and Insights from Cryptocurrency Lab*, [online] 2015. [cit. 13.01.2022]. Available at: http://eprint.iacr.org/2015/460

[8] NAKAMOTO, Satoshi. Bitcoin: Peer-to-Peer Electronic Cash System, [online]. 2008. [cit. 13.01.2022]. Available at: https://bitcoin.org/bitcoin.pdf

[9] SZABO, Nick. *Idea of Smart Contracts* [online]. c 1997 [cit. 13. 1. 2022]. Available at: https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html

[10] OLICKEL, Hrishi. Why Smart Contracts Fail: Undiscovered bugs and what we can do about them [online]. *Medium*. 9. 9. 2019 [cit. 13. 1. 2022]. Available at: https://hrishiolickel.-medium.com/why-smart-contracts-fail-undiscovered-bugs-and-what-we-can-do-about-them-119aa2843007

[11] RASKIN, Max. Law And Legality Of Smart Contracts. *Georgetown Law Technology Review* [online]. 2017, p. 305-340 [cit. 13.01.2022]. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959166

[12] SEGAL, David. My Puzzling Entry in Crypto World. *New York Times*, [online]. 2021. [cit. 13.01.2022]. Available at: https://www.nytimes.com/2021/08/17/insider/cryptocurrency-hype-coin.html

[13] SKLAROFF, Jeremy. *Smart Contracts and Cost of Inflexibility*. Rochester, NY: Social Science Research Network, [online] 2017, p. 264-302. [cit. 13.01.2022]. Available at: https://papers.ssrn.com/abstract=3008899

[14] State of DApps — What's DApp [online]. c 2022. [cit. 13. 1. 2022]. Available at: https://www.stateofthedapps.com/whats-a-dapp

[15] ZINOVYEVA, Elizaveta,REULE, Raphael C.G., HÄRDLE, Wolfgang K. *Understanding Smart Contracts: Hype or Hope?*. Rochester, NY: Social Science Research Network, [online]. 2021. [cit. 13.01.2022]. Available at: https://papers.ssrn.com/abstract=3804861

PROBLEMS WITH ALGORITHMIC CONTENT MODERATION IN SOCIAL NETWORKS[1]

ROBERTA HULANSKÁ[2]

1. INTRODUCTION

“*During the past few years, the global conversation about responsible technology has intensified. Increasingly, we are acknowledging that technology is not and can never be neutral, that it holds significant implications for people and society, and that intelligent technologies have consequences that can disenfranchise or target vulnerable populations.“*[3] Part of the technologies is algorithms. They increasingly dominate many aspects of modern society. Algorithms affect our lives in every possible way, with serious and significant impacts. A field that is considered to be very influenced by automated decision-making is social media. These platforms do much more than passively distribute user content and facilitate user interactions. They now have near-total control of users’ online experience and content moderation.[4] The US Supreme Court has affirmed the importance of social media platforms as venues for free speech in Packingham v North Carolina. In giving the lead judgment,

[1] Esej byla zpracována v semestru podzim 2021 v rámci předmětu MVV1368K Privacy and Personal Data na téma Personal data protection online III – Automatic decision making. / The essay was written in the autumn 2021 semester for the course MVV1368K Privacy and Personal Data on the topic of Personal data protection online III – Automatic decision making

[2] Bc. Roberta Hulanská je studentkou magisterského studijního programu Právo a právní věda na Právnické fakultě Masarykovy univerzity, kontakt: 471219@mail.muni.cz

[3] ETLINEGR, Susan. What’s So Difficult about Social Media Platform Governance? *Centre for International Governance Innovation*, [online]. 2019. [cit. 18. 11. 2021], p. 21.

[4] CASTETS-RENARD, Céline. Algorithmic content moderation on social media in EU law: illusion of perfect enforcement. *University of Illinois Journal of Law, Technology & Policy*, [online]. 2020, n. 2, [cit. 18.11. 2021], p. 283.

Justice Kennedy explained that 'these websites can provide perhaps the most powerful mechanisms available to a private citizen to make his or her voice heard.[5]

Unfortunately, there are various negative aspects related to free speech on social media platforms. Hate speech, fake news, and content inciting violence have become the unfortunate norm. Because of this, nowadays, platforms are required to moderate content, mainly remove illegal content. But the content moderation does not work like in the past when platforms or forums were managed by administrators (humans). Today, big platforms like Facebook, Twitter or Google use algorithmic decision-making that helps scale down the massive task of content moderation. It seems like a very effective tool that provides perfect enforcement.[6] But it is not that simple. The problem comes when deciding how the algorithm will work in order to tackle content. The comprehensive enforcement of policy violations largely depends on the manner in which companies choose to search, detect, and review potentially violative content. Despite the vast improvements in technology and the evolution of social media, the algorithmic content moderation method is still far from perfect.[7]

Content moderation and distribution — in other words, the composition of users' feeds and the accessibility and visibility of content on social media — happen through a combination of human and algorithmic decision-making processes.[8] In this essay, I will focus on algorithmic processes and point out some of the problems that arise when it comes to algorithmic content

[5] PACKINGHAM v. NORTH CAROLINA, 582 U.S. *Justia US Supreme Court*, [online]. 2017. [cit. 18. 11. 2021]. Available at: https://supreme.justia.com/cases/federal/us/582/15-1194/

[6] CASTETS-RENARD, Céline. Algorithmic content moderation on social media in EU law: illusion of perfect enforcement. *University of Illinois Journal of Law, Technology & Policy*, [online]. 2020, n. 2, [cit. 18. 11. 2021], p. 283.

[7] YOUNG, Greyson. K. How much is too much: the difficulties of social media content moderation. *Information & Communications Technology Law*, [online]. 2021. [cit. 18. 11. 2021], p. 4.

[8] DOCQUIR, Pierre F. The Social Media Council: Bringing Human Rights Standards to Content. *Centre for International Governance Innovation*, [online]. 2019. [cit. 18. 11.2 021], p. 9.

moderation in social networks. Firstly, the author will briefly introduce terminology, and then open the topic of relevant problems.

2. DEFINITION OF ALGORITHMIC CONTENT MODERATION IN THE CONTEXT OF SOCIAL MEDIA NETWORKS

Algorithmic moderation can be defined in various ways. One, broad, definition is provided by Grimmelmann, who characterizes it as the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse. In Grimmelmann's understanding, moderation includes not only the administrators or moderators with the power to remove content or exclude users but also the design decisions that organise how the members of a community engage with one another.[9]

The narrower definition is provided by the authors Gorwa, Binns and Katzenbach. They define it as systems that classify user-generated content based on either matching or prediction, leading to a decision and governance outcome (e.g. removal, geoblocking or account takedown). Algorithmic content moderation involves a range of techniques from statistics and computer science, which vary in complexity and effectiveness. They all aim to identify, match, predict, or classify some piece of content on the basis of its properties or general features.[10]

The content moderation process at social media companies can be broken down into three distinct stages: creation, enforcement and response. Creation describes the development of the rules (the terms and conditions) that platforms use to govern user conduct. Enforcement entails the flagging of content as problematic, the decision on whether the content is in breach of the terms and conditions, and what actions should be taken. Response, the final stage, describes the internal appeals process used by platforms and

[9] GRIMMELMANN, James, The virtues of moderation. *Yale Journal of Law & Technology*, [online]. 2015, n. 17. [cit. 18. 11. 2021], p. 42.

[10] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021].

the methods of collective action activists might use to change the platform from the outside.[11]

3. SELECTED PROBLEMS WITH ALGORITHMIC CONTENT MODERATION IN SOCIAL NETWORKS

In the past decades, we have experienced dramatic advancements in technology and we have seen a massive growth of social media platforms. Nowadays, social media platforms are heavily increasing their use of artificial intelligence to moderate content posted by users. Using algorithms to find and remove violative content from users' newsfeeds takes an ex-ante approach to moderation. Algorithms aim to apply the platforms' policies to content as it is uploaded to the site and remove prohibited materials before other users are able to see them. Although using automatic moderation systems may prevent prohibited content from impacting or influencing many users, this method has many problems, for instance, a lack of understanding of a post's intention, context, or idiom.[12]

The basic problem, when it comes to algorithmic content moderation in social platforms, is that these systems cannot tackle all issues that are needed. *„When it comes to content moderation, AI programs are not adept at understanding context and nuance, so they make mistakes that can result in "false positives" (flagging an innocuous video, statement or photo) or "false negatives" (missing a violent or otherwise undesirable post). In the world of social media, false positives prompt protests over censorship, for example, when a platform removes a post by an organization that is sharing it to raise awareness of a human rights violation, while false negatives expose the company to legal liability, if, say, it fails to recognize and remove prohibited content within a stipulated time period.“*[13] Language is another problem that is linked to this. While language technology continues to improve rapidly, it remains highly depend-

[11] COMMON, MacKenzie. Fear the Reaper: how content moderation rules are enforced on social media. *International Review of Law, Computers & Technolog*, [online]. 2020, vol. 34, n. 2. [cit. 18. 11. 2021], p. 127.

[12] YOUNG, Greyson. K. How much is too much: the difficulties of social media content moderation. *Information & Communications Technology Law*, [online]. 2021. [cit. 18.11.2021], p. 9.

ent on high volumes of labelled and clean data to achieve an acceptable level of accuracy.[14]

3.1 TRANSPARENCY

A common critique of automated decision-making is the lack of transparency. Content moderation has been a secretive process. Years of pressure by researchers, journalists and activists have recently led to notable efforts by companies (e.g. Facebook) to make their moderation practices more transparent (publication of the 'Community Standards' could be named as an example). However, it is still not enough plus the rapid push toward algorithmic moderation in the past few years threatens to reverse much of this progress.[15]

Although total transparency cannot be expected, minimum standards of decisional transparency are essential to allow both ordinary users and critical experts to understand the patterns of governance within which they are embedded.[16]

According to Van Dijck, transparency is not a reciprocal action on social media but rather surprisingly one-sided. "*Users are increasingly encouraged to share as much as possible on social media platforms, an action that not only populates the platform with original content but also provides valuable data that can be sold to third-party advertisers.*[17] *Meanwhile, social media companies continue to perform the proverbial dance of the seven veils, obscuring their actions in code and proprietary arguments, thus pre-empting attempts to hold them accountable.*"[18]

[13] ETLINEGR, Susan. What's So Difficult about Social Media Platform Governance? *Centre for International Governance Innovation*, [online]. 2019. [cit. 18. 11. 2021], p. 22.

[14] Ibidem. p. 23.

[15] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021].

[16] Ibidem.

[17] KAUN, Anne. Jose van Dijck: Culture of Connectivity: A Critical History of Social Media. Oxford: Oxford University Press. 2013. *MedieKultur: Journal of media and communication research*. [online]. 2014, vol. 30. [cit. 18.11.2021]. p. 61.

3.2 FAIRNESS

Technology is not neutral but instead embedded with values and politics. Recent years have seen substantial discussion about the potential for algorithmic decision-making systems to have unfair or discriminatory impacts on different groups, such as protected classes under anti-discrimination law. Content classifiers in general, whether used for recommendation, ranking, or blocking, may be more or less favourable to content associated with gender, race and other protected categories, and thus entrench forms of representational harm against such groups.[19] There is a consensus among international experts on freedom of expression that the mere regulation of speech by contract fails to provide adequate transparency and protection for freedom of expression and other human rights. Individual users have little or no remedy to address content removal and they are given no guarantee for the protection of individual freedoms.[20]

Even a perfectly 'accurate' toxic speech classifier will have unequal impacts on different populations because it will inevitably have to privilege certain formalisations of an offence above others, disproportionately blocking (or allowing) content produced by (or targeted at) certain groups. For instance, hate speech classifiers designed to detect violations of a platform's guidelines could be disproportionally flagging language used by a certain social group, thus making that group's expression more likely to be removed.[21]

[18] COMMON, MacKenzie. Fear the Reaper: how content moderation rules are enforced on social media. *International Review of Law, Computers & Technology*, [online]. 2020, vol. 34, n. 2. [cit. 18. 11. 2021], p. 141-142.

[19] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021].

[20] DOCQUIR, Pierre F. The Social Media Council: Bringing Human Rights Standards to Content. *Centre for International Governance Innovation*, [online]. 2019. [cit. 18. 11.2 021], p. 10.

[21] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021].

Another important variable to keep in mind is bias and cultural issues for human moderators. This essay's goal is not to talk about bias and the human factor, but I wanted to emphasize that it is important to know who exactly creates the algorithms that are used for content moderation. One of the causes of human rights violations occurrences on social media platforms is that limitations on expression are applied inconsistently and may replicate the biases experienced by the predominantly white and male staffers at social media platforms who devise content assessment strategies.[22]

4. SELECTED SPECIFIC PROBLEMS

4.1 TOXIC SPEECH AND HARASSMENT

Harassment has long been an issue in online spaces, particularly gender-based harassment, which is prevalent across many online platforms. According to a survey executed in 2014, 73% of adult American internet users had witnessed harassment online and 40% had personally been harassed.[23]

Any platform that enables the communication between users faces problems of potentially offensive speech, personal attacks and abuse that could harm users, distort conversation or even drive certain contributors away.[24] Because of this, there have been efforts by several social media platforms to build programs that will find these types of text.

In the past few years, Facebook has responded to growing pressure around hate speech (especially from EU member states) by developing classifiers that are trained to predict whether text may constitute hate speech, and based on that score, flag it for human review. Instagram and YouTube as well have started tackling this issue by developing toxic speech

[22] Ibidem.

[23] GEIGER, R. Stuart. Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Communication & Society*. [online]. 2016. vol. 19, n. 6. [cit. 18. 11. 2021], p. 787.

[24] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18.11.2021].

classifiers to identify certain types of comments.[25] On the other hand, Twitter, Inc. has generally taken a far more hands-off approach to moderation than other social networking sites, and the design of the platform affords unsolicited interactions in ways that others do not.[26]

It would not be far from the truth to say that it is virtually impossible to curb this type of post. The clearest problem is the language - it is incredibly complicated, personal and context-dependent: even words that are widely accepted to be slurs may be used by members of a group to reclaim certain terms. For instance, there was a research collaboration between Google and the Wikimedia Foundation regarding algorithmic moderation of toxic speech and the results were quite surprising. For example, the single-term comment ‘Arabs’ was classed as 63% toxic, while the phrase ‘I love führer’ was only 3% toxic.[27]

4.2 TERRORISM

In 2017, Google, Facebook, Twitter and Microsoft announced the creation of the GIFCT. This organisation remains highly secretive, has a board made of ‘senior representatives from the four founding companies and publishes little about its operations. However, the organisation has been particularly focused on the improvement of automated systems to remove extremist images, videos and text.[28]

Even though it remains unknown how these systems really function, we know they are not 100% effective based on numerous examples. For instance, more platforms have traditionally allowed terrorist images if they are being used by a reputable news organisation or in order to express disapproval or condemnation of a group. However, automated systems re-

[25] Ibidem.

[26] GEIGER, R. Stuart. Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Communication & Society*. [online]. 2016. vol. 19, n. 6. [cit. 18. 11. 2021], p. 788.

[27] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021].

[28] Ibidem.

moved thousands of videos that had been uploaded to YouTube by civil society groups and activists to document atrocities conducted during the Syrian Civil War.[29] Machine learning systems are poor at making such difficult context-dependent judgements.

The platforms were used to livestream the terror attacks in Christchurch, New Zealand. They have also been used as a tool for ethnic cleansing in Myanmar.[30] This is a disturbing problem. These videos are unpredictable, difficult to interrupt, and are not subject to algorithmic moderation because the content is simultaneously shared and uploaded to the platform.[31] Algorithmic moderation systems cannot tackle them very well.

5. CONCLUSION

Critical conversations about algorithmic moderation systems often emphasise the challenges that these systems face nowadays. It is commonly pointed out that it is very difficult for predictive classifiers to make difficult, contextual decisions on slippery concepts like hate speech for instance, and that automated systems at scale are likely to make hundreds, if not thousands, of incorrect decisions on a daily basis.[32] Even though there are also positives arising from the usage of algorithmic content moderation on social media platforms, the purpose of this essay was to briefly comment on some of the most debated problems.

The most important problem of algorithmic content moderation is that these systems cannot tackle all issues that are needed. The world and its communities are so complex that it is just not possible. From this basic

[29] BROWNE, Malachy. YouTube Removes Videos Showing Atrocities in Syria. *The New York Times*, [online]. 2017. [cit. 18.11.2021]. Available at: https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html

[30] ETLINEGR, Susan. What's So Difficult about Social Media Platform Governance? *Centre for International Governance Innovation*, [online]. 2019. [cit. 18. 11. 2021], p. 21.

[31] COMMON, MacKenzie. Fear the Reaper: how content moderation rules are enforced on social media. *International Review of Law, Computers & Technology*, [online]. 2020, vol. 34, n. 2, [cit. 18. 11. 2021], p. 131.

[32] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021].

problem, others arise, namely challenges regarding transparency and fairness. Maybe calling them goals would be more fitting, as it is again impossible to reach total transparency and fairness, online and offline, as well. Hopefully, we will see some improvement in this area. Furthermore, social media platforms have to tackle hate speech and terrorism among others every day.

I wish I could finish this essay by saying that these are all the problems. There are, unfortunately, many more. It is not even possible to say what works and what does not when a question about how to solve the problems mentioned in this essay would come up. Addressing these issues is not as straightforward as it seems. In addition to the legal, social and cultural dynamics at play, there are other factors we must consider: the scale of social media platforms, the technologies on which they are built and the economic environments in which they operate.[33]

It is unlikely that social media is ever going to be given a perfect solution for how to handle content moderation. A platform's terms of use need to be specific enough to capture and remove posts that need to be deleted and not remove the ones that are not problematic, but broad enough in order to include every unsuited content. Maybe rather than try to blame the platforms for not doing enough in this department, we should think about the core of this problem - the people, the users of social media.

6. BIBLIOGRAPHY

[1] BROWNE, Malachy. YouTube Removes Videos Showing Atrocities in Syria. *The New York Times*, [online]. 2017. [cit. 18.11.2021]. Available at: https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html

[2] CASTETS-RENARD, Céline. Algorithmic content moderation on social media in EU law: illusion of perfect enforcement. *University of Illinois Journal of Law, Technology & Policy*, [online]. 2020, n. 2, p. 283-324 [cit. 18. 11. 2021]. Available at: https://heinonline.org/HOL/P?h=hein.journals/jltp2020&i=295

[33] ETLINEGR, Susan. What's So Difficult about Social Media Platform Governance? *Centre for International Governance Innovation*, [online]. 2019. [cit. 18. 11. 2021], p. 22.

[3] COMMON, MacKenzie. Fear the Reaper: how content moderation rules are enforced on social media. *International Review of Law, Computers & Technology*, [online]. 2020, vol. 34, n. 2, p. 126-152. [cit. 18. 11. 2021]. Available at: https://doi.org/10.1080/13600869.2020.1733762

[4] DOCQUIR, Pierre F. The Social Media Council: Bringing Human Rights Standards to Content. *Centre for International Governance Innovation*, [online]. 2019. p. 9-12. [cit. 18. 11. 2021]. Available at: https://www.jstor.org/stable/resrep26127.4

[5] ETLINEGR, Susan. What's So Difficult about Social Media Platform Governance? *Centre for International Governance Innovation*, [online]. 2019, p. 20-26, [cit. 18. 11. 2021]. Available at: https://www.jstor.org/stable/resrep26127.6

[6] GEIGER, R. Stuart. Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Communication & Society*. [online]. 2016. vol. 19, n. 6, p. 787-803. [cit. 18. 11. 2021]. Available at: https://doi.org/10.1080/1369118X.2016.1153700

[7] GORWA, Robert, BINNS Reuben and KATZENBACH Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [online]. 2020, n. 7. [cit. 18. 11. 2021]. Available at: https://journals.sagepub.com/doi/full/10.1177/2053951719897945

[8] GRIMMELMANN, James. The virtues of moderation. *Yale Journal of Law & Technology*, [online]. 2015, n. 17, p. 42-109. [cit. 18. 11. 2021]. Available at: https://digitalcommons.law.yale.edu/yjolt/vol17/iss1/2/

[9] PACKINGHAM v. NORTH CAROLINA, 582 U.S. *Justia US Supreme Court*, [online]. 2017. [cit. 18. 11. 2021]. Available at: https://supreme.justia.com/cases/federal/us/582/15-1194/

[10] KAUN, Anne. Jose van Dijck: Culture of Connectivity: A Critical History of Social Media. Oxford: Oxford University Press. 2013. *MedieKultur: Journal of media and communication research*. [online]. 2014, vol. 30, p. 3. [cit. 18. 11. 2021]. Available at: DOI: 10.7146/mediekultur.v30i56.16314

[11] *YOUNG, Greyson. K. How much is too much: the difficulties of social media content moderation. Information & Communications Technology Law, [online]. 2021, p. 1-16. [cit. 18. 11. 2021]. Available at: https://doi.org/10.1080/13600834.2021.1905593*

EU TAKING THE EASIER PATH TO REGULATE AI[1]

TENA KRZNARIĆ[2]

1. INTRODUCTION

Technology triggers social and economic progress. It is difficult to develop it and even more difficult to control it. When it comes to the regulation of technology there are two points of view. The first one is represented by lawyers as laymen and the second one by engineers. Lawyers tend to see things as potential abuse ground and danger, while engineers are turned to progress and achieving the greatest potential technologies can offer us. Law will never be able to predict every situation. What we regulate today, most likely will barely be usable in the future. EU's attempt to regulate the idea of AI is a nice try of putting everything that we have achieved together while bringing bureaucratisation of innovation which is not respected from the innovator's standpoint. Seems like the EU approaches the new ideas by telling them "Yes, but…" and turning on the danger alarm.

2. SUBLIMINAL BEHAVIOUR MANIPULATION

In Article 5 (1) (a) of the Proposal[3] EU has explicitly stated concern about AI systems using subliminal techniques. Subliminal techniques arise from

[1] Esej byla zpracována v semestru podzim 2021 v rámci předmětu MVV57917K Regulating Disruptive Technologies na téma Artificial Intelligence. / The essay was written in the autumn 2021 semester for the course MVV57917K Regulating Disruptive Technologies on the topic of Artificial Inteligence

[2] Tena Krznarić je studentkou Faculty of Law, University of Zagreb, kontakt: tena.krznaric.pravo@gmail.com

[3] Proposal for a regulation of the European parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence act) and amending certain Union legislative acts [online]. 2021. [cit. 02. 12. 2021]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

the term “subliminal perception” whose starting point is a thought that “*it is possible to influence human thoughts, feelings, and behaviours through various stimuli without the conscious knowledge of the person to whom is affected.*”[4] The core of this prohibition is manipulation. The biggest problem arising from this provision is that it is too general and too abstract. The only clear part is the intention of the EU to prevent physical and psychological harm to individuals. Namely, which subliminal techniques are those that cause physical and psychological harm? Psychology recognises two types of subliminal techniques: visual and audio.[5] However, we are still not familiar with techniques used by AI systems that can cause such described harm. Concerns about subliminal techniques are not new, but mainly affect the area of commercial activities. For example, the Croatian Act on Electronic Media forbids the usage of subliminal techniques in audio-visual commercial communication.[6] Although noticed by many researchers, this ban does not apply to commercial practices[7] which are covered by Unfair Commercial Practices Directive.[8] On the other hand, some connect it to “dark patterns”.[9] There are many definitions of dark patterns, still non-official, though for a better understanding of the term this one is used: “Dark patterns are user interfaces whose designers knowingly confuse users, make

[4] MILIŠA, Zlatko and NIKOLIĆ, Gabrijela. Subliminalne poruke i tehnike u medijima. *Nova prisutnost: časopis za intelektualna i duhovna pitanja*. Kršćanski akademski krug (KRAK), [online]. 2013, vol. XI, issue 2, [cit. 02.12. 2021], p. 297.

[5] Ibid., p. 298.

[6] Act on Electronic Media; NN 111/21; Art. 21 (3), [online]. [b.r.]. [cit. 02. 12. 2021]. Available at: https://www.zakon.hr/z/196/Zakon-o-elektroni%C4%8Dkimmedijima

[7] VEALE, Michael and BORGESIUS, Frederik Zuiderveen. Demystifying the Draft EU Artificial Intelligence Act.[online]. c 2022, [cit. 02. 12. 2022], p. 98-100.

[8] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (‘Unfair Commercial Practices Directive’), [online]. 2005. [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/dir/2005/29/oj/eng

[9] PROPP, Kenneth and MACCARTHY, Mark. *Machines learn that Brussels writes the rules: The EU’s new AI regulation* [online]. 2021. [cit. 02. 12. 2021]. Available at: https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/

it difficult for users to express their actual preferences, or manipulate users into taking certain actions.”[10] Norwegian Consumer Council engaged in studying this topic defined five categories of dark patterns in digital services such as: default settings, ease, framing, rewards and punishment, and forced action.[11] Regarding our topic one interesting comment was given by the Council: “none of these categories of nudging is inherently unethical, and can conceivably be used to achieve results that are in the users’ best interests.“[12] Luguri and Strahilevitz conducted two experiments by using dark patterns. The first goal was to see to which extent they affect people’s decisions and secondly, do all dark patterns affect people’s decisions evenly or do some affect them more. In the first experiment, they distinguished mild and aggressive dark patterns while “selling” protection from identity theft services. The result showed that, with the usage of mild dark patterns, sales increased double and with aggressive, it quadrupled. They concluded that the law should regulate the subtle use of dark patterns due to their ability to affect more vulnerable groups.[13] In the second experiment, they distinguished dark patterns that affected the decision-making of purchasing the service and those which had no effect.[14] Those affected the most were hidden information, obstruction, trick questions, and social proof.[15] It seems like the greatest concern of the EU should not be physical or psychological harm due to more damage to individuals occurring in the economical or privacy area. Those are covered under Commercial Practices

[10] LUGURI, Jamie and STRAHILEVITZ, Lior. *Shining a Light on Dark Patterns*. Rochester, NY: Social Science Research Network, [online]. 2021. [cit. 02. 12. 2021]. DOI: 10.2139/ssrn.3431205

[11] Norwegian Consumer Council; *DECEIVED BY DESIGN- How tech companies use dark patterns to discourage us from exercising our rights to privacy*. [online]. 2018. [cit. 02. 12. 2021]. p. 12.

[12] Ibid.

[13] LUGURI, Jamie and STRAHILEVITZ, Lior. *Shining a Light on Dark Patterns*. Rochester, NY: Social Science Research Network, [online]. 2021. [cit. 02. 12. 2021], p. 46-47.

[14] Ibid.

[15] Dark patterns affecting the most: “Hidden information (smaller print in a less visually prominent location), obstruction (making users jump through unnecessary hoops to reject a service), trick questions (intentionally confusing prompts), and social proof (efforts to generate a bandwagon effect).”

Directive and GDPR[16] and should be extended to AI systems. Subliminal techniques do not affect human behaviour for a longer period, just one second and longer only if pointed out and individuals process the given information.[17] In order to prohibit such AI systems, the EU should extend goals regarding the protection and specify forbidden techniques. Though, not to sabotage itself conduct approach to regulation would serve better.

3. EXPLOITATIVE BEHAVIOUR MANIPULATION

To understand the basic idea behind Article 5 (1) (b) we have to understand how AI can exploit human behaviour. The thing is, it is only theoretical but the research has to start somewhere. CSIRO's Data61[18] made a study on how AI can be used to influence human decision-making by exploiting vulnerabilities in an individual's habits and patterns. Three experiments were conducted in which people played games against a computer. In the first one participants had to choose between squares on the screen in order to achieve an award. The AI was learning their choice patterns and guided them to their choice with a success rate of 70 %.[19] In the second one, participants had to press a button every time a certain shape appears on the screen. The AI started to arrange the sequence of symbols which resulted in a 25 % increase in mistakes made by participants.[20] The third one was more complex. In this one the participant gained the role of an investor who had to invest in a trustee and the AI played the role of a trustee.

[16] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [online]. 2016 [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng

[17] RUCH, Simon, ZÜST, Marc Alain and HENKE, Katharina. Subliminal messages exert long-term effects on decision-making. *Neuroscience of Consciousness* [online]. 2016, vol. 2016, issue 1, [cit. 02. 12. 2021], p. 5.

[18] Data and digital specialist arm of Commonwealth Scientific and Industrial Research Organisation

[19] DEZFOULI, Amir, NOCK, Richard and DAYAN, Peter. Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences,* [online]. 2020, vol. 117, issue 46, [cit. 02. 12. 2021], p. 29223.

[20] Ibid., p. 29224.

The game was played in several rounds and two modes. After every round, the AI had to return some amount of money to the participant and the participant had to decide how much he wants to invest in the next round. The amount of money returned to the participant was depending on the mode. While playing in the first mode the AI was trying to maximize its profit and in the second one, the AI sought to distribute profit fairly between itself and the participant investing. The experiment showed the success of AI in gaining profit in both modes.[21] The purpose of these experiments was for the AI to learn from human actions and to seek and target their vulnerabilities. This finding confirms the EU's fear of certain groups and makes it reasonable. However, it must be emphasized that such use of AI does not necessarily bring harm, because AI's learning process, has the possibility to alert the user on his/her vulnerabilities and guide them to better decisions. Technology itself is not a problem but a creator and his intention behind it are. This leads to a conclusion that it is not necessary to ban this type of AI, moreover, a way of using it should be regulated because the creator can set up parameters to achieve desired behaviour of the subject, which means the intention of one setting it up leads to harm to the individual. While the subject of the study is still its impact, the ban of such AI systems seems like a premature decision based on fear, but it puts emphasis on what needs to be monitored.

4. SOCIAL SCORING

Article 5(1) (c) can reasonably be justified but it is yet to be seen in which direction. One of the first points we need to pay attention to is the possibility to use AI systems in creating and conducting social scoring systems. What are social scoring systems? Social scoring tends to collect data of every individual which doesn't include just regularly collected personal data such as name, surname, address, work position, etc., but also collects data on individuals' psychological and physical characteristics.[22] The aim of

[21] Ibid., p. 29225.

[22] Kapersky daily, Social scoring systems: current state and potential future implications. In: *Kasperksy daily* [online]. c 2021 [cit. 02. 12. 2021]. Available at: https://www.kaspersky.com/blog/social-scoring-systems/

collecting such data is to make rankings among people. We can literally imagine people having a cloud above their heads that shows the number of credits they have, remembering that the number is not constant but variable depending on how that person behaves and what they do. The simplest example would be on with whom they are friends, how much are they engaged in studying or working, do they contribute to charity. It can also include their (non)healthy habits or even how emotionally satisfied they are regarding their work, surroundings, or which political ideology they gravitate to. Not to forget examples like paying bills on time or repaying loans. What actually is a problem in regards to social scoring is who uses it, how it's being used, and why. The first problem, to which Article 5(1) (c) refers to, is using AI in order to create a social scoring system by public authorities. China already started to use it years ago. As already mentioned a person gains or loses points based on what they do, as in China everyone started with basic 1000 points. In regards to the outcome of counting points, individuals whose behaviour leads to the constant loss of points in China are in a so-called position of being a blacklisted person.[23] In other words, people who lose points are deprived of some rights such as access to public authorities, a ban on travel, and many more. In China not only public authority uses it but also private companies. The most tremendous use of it is by public authorities due to the fact that their basic purpose is to resolve social problems to service society. This does not include the use of removing blacklisted persons from society.[24] Another country that introduced this system is the United Kingdom, and similarly to the Chinese implementation, they intended to introduce rewards and punishments depending on the score. Several studies conducted among UK people gained insight into what they were expecting from it in comparison to the Chinese system. When it comes to rewards the most appealing were

[23] NAST, Condé. The complicated truth about China's social credit system. *Wired UK* [online]. 2019. [cit. 02. 12. 2021]. ISSN 1357-0978. Available at: https://www.wired.co.uk/article/china-social-credit-system-explained

[24] CANALES, Katie. China's „social credit" system ranks citizens and punishes them with throttled internet speeds and flight bans if the Communist Party deems them untrustworthy. In: *Business Insider* [online]. 2021. [cit. 02. 12. 2021]. Available at: https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4

in the following priority order: healthcare, lower energy bills, favourable interest rates on loans, better schooling for children, and travel privileges. On the other side, some penalties were introduced- public naming and shaming, denied access to credit cards, and lack of possibility to apply for certain jobs.[25] In a democratic society use of this form of punishment and reward system cannot be reasonable and is threatening to democratic values and human rights. In this day and age when modern contemporary society developed certain values, which lacked in previous centuries, it cannot be allowed that basic democratic values such as equality (especially before the law), freedom of decision-making and speech, social justice, and others are taken away. Most rights and freedoms would be greatly endangered by the implementation of social scoring in the presented way by the side of public authorities.

Attention must be paid to whether there is a need to use this type of technology. Modern society can simplify and accelerate some of the decision-making processes with it. For example, it can help with the decision-making process for granting loans, so in the case of several loans, it can count points on the financial history of a person, or when applying for a job it can collect necessary data which can ease the process of selecting a proper candidate and eliminating those who don't meet requirements.

In the conclusion to point (c) of Article 5, it can be noticed that there is a need to regulate this form of using AI systems. As it states in the article, with direct interpretation, the use of these technologies with intention of unjustified social scoring in everyday use is only prohibited when it intends to create detrimental or unfavourable treatment of individuals or groups based on their social behaviour, unrelated to the context why was originally collected for. In other words, the intention of Article 5 (c) is not to completely ban the use of AI systems in creating social scoring but to prohibit certain use by public authorities which threaten modern society, rights, and freedoms for which society was fighting for a really long time. The only negative aspect found in point (c) would be that it is only

[25] ABC Finance. Surviving The Social Credit Score. In: *ABC Finance* [online]. c 2021 [cit. 02. 12. 2021]. Available at: https://abcfinance.co.uk/blog/surviving-the-social-credit-score/

orientated toward public authorities and does not take private companies into account.[26] Even though public authorities must pay more attention to respecting human rights and freedoms, it doesn't mean that private companies or similar subjects are excluded from it. They do have the freedom to conduct their business in the way they want to as long as they do it with respect and in conformity with international treaties, the constitution, and other legal acts. We can see that trends in the world do represent a breach of some basic human rights and it is necessary to control it with respect to its prohibition of certain unfair practices is the best solution as long as it concerns only the unfair practice. Point (c) openly limits the prohibition of the results that are not compatible with a democratic society and the general explanation of these results leaves enough room to determine in each specific case if there was that kind of intention. The best result this prohibition can have is the prevention of direct discrimination in exercising at least basic rights given by the state authorities.

5. REAL-TIME REMOTE BIOMETRIC ID

Firstly, detecting the problem of the scope and aim of Article 5 (1) (d) can start from the current world situation, and by that is specifically meant the COVID-19 pandemic. This situation can be seen as relevant for point (d) (III). Why so? Namely, as the provision states 'real-time' remote biometric identification systems in publicly accessible space can be used for detection, localisation, identification, or persecution of a perpetrator or suspect of a criminal offence but with setting out limitations. For possible use criminal offence has to be included in Article 2 (2) of Council Framework Decision 2002/584/JHA 62[27] and punishable in the Member State by a custodial sentence or detention for a maximum period of at least three years. So where is the problem with the COVID-19 pandemic? In the aim of pre-

[26] EBERS, Martin et al. The European Commission's Proposal for an Artificial Intelligence Act —A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J*. Multidisciplinary Digital Publishing Institute, [online]. 2021, vol. 4, issue 4, [cit. 02. 12. 2021], p. 592.

[27] Council Framework Decision of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States (2002/584/JHA) [online]. 2009. [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/dec_framw/2002/584/2009-03-28/eng

vention of infectious diseases states include a measure of quarantine or self-isolation prescribed under the law. Another crucial thing is that national law also contains provisions on spreading infectious diseases in national criminal law acts. In regard to this problem, we can turn to Croatian national law. As mentioned, the measure of quarantine or self-isolation in Croatian law is prescribed in the Act on Protection of the Population from Infectious Diseases.[28] When it comes to criminal law Croatian Criminal Law Act[29] in Chapter XIX Article 180 contains a provision concerning the spread and transmission of infectious diseases. The problem here lies in the fact that neither Article 2 (2) of the Council Framework Decision contains this offence nor does the Croatian Criminal Law Act predict imprisonment for a maximum period of at least three years. Why is it at all that important? Because use of a "real-time" biometric identification system in publicly accessible spaces can be helpful in detecting suspects of criminal offences which in this case would be any person to whom a measure of quarantine or self-isolation was prescribed for a specific period of time. This is an example of the positive use of a biometric system, which under Article 5 would be declined due to the fact that none of the exemptions includes the protection of health as a reason. A similar practice has been seen in Slovakia with the eQuarantine app which is based on biometrics.[30] The app was made by a Slovak company Innovatrics and it is based in the EU. The potential problem which was detected was personal data collection due to the fact that it wasn't clear who will collect data, how will be stored, and for how long.[31] From the side of privacy and personal data this can be seen as a problem that initiated the creation of Article 5 however, it can have

[28] Act on Protection of the Population from Infectious Diseases, NN 79/07 , 113/08 , 43/09 , 130/17 , 114/18 , 47/20 , 134/20, [online]. [b.r.]. [cit. 02. 12. 2021]. Available at: https://www.zakon.hr/z/1067/Zakon-o-za%C5%A1titi-pu%C4%8Danstva-odzaraznih-bolesti

[29] Criminal Law Act, NN 125/11, 144/12, 56/15, 61/15, 101/17, 118/18, 126/19, 84/21, [online]. [b.r.]. [cit. 02. 12. 2021]. Available at: https://www.zakon.hr/z/98/Kazneni-zakon

[30] Innovatrics. Self-Isolation Rather than Quarantine - Thanks to Face Recognition. In: *Innovatrics* [online]. 18. 6. 2020. [cit. 02. 12. 2021]. Available at: https://www.innovatrics.com/news/covid19-self-isolation-rather-than-quarantine-thanks-to-face-recognition/

positive use globally so it can help to stop the spread of disease. What can be done in order to have a useful tool but still protect data? More control over its use is needed, a detailed explanation of data that has to be used needs to be prepared and the question of data storage and collection needs to be transparent. This can all be resolved by following the rule prescribed in GDPR.[32] So what do we get with Article 5 prohibition? The prohibition actually doesn't serve its purpose. The prohibition in regard to what was said is too broad.[33] A better solution would be to exercise more control over the use and not prohibit it. Article 5 (d) exceptions are too narrow and in the future, with all the development there will surely be a need to broaden it. If a focus is put on a social purpose we can also refer to the limitation of rights. In the eyes of some, using biometric identification systems can be seen as a violation of the right to privacy. On the other hand, here we have a good example of where a test of proportionality can be used. If a state wants to improve public security, especially when it comes to infectious diseases or some other aspects of security, and in order to protect public health and wants to use this type of technology. Depending on the main aim we can see that in order to protect public health there is room for some limitations of the right to privacy. People cannot reasonably expect privacy and at the same time put in danger a larger number of people. Only the methods and procedure need to be transparent and control over it has to be exercised. Another key thing to point out is the fact that this wide prohibition has negative consequences for EU companies because it limits technology development. Also, a commercial component is in danger be-

[31] SIROTNIKOVA, Miroslava German. *Question Marks over Slovak Quarantine App Fuel Privacy Concerns* [online]. 2020. [cit. 02. 12. 2021]. Available at: https://balkaninsight.com/2020/05/20/question-marks-over-slovak-quarantine-app-fuel-privacy-concerns/

[32] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [online]. 2016. [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng

[33] EBERS, Martin et al. The European Commission's Proposal for an Artificial Intelligence Act —A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J*. Multidisciplinary Digital Publishing Institute, [online]. 2021, vol. 4, issue 4, [cit. 02. 12. 2021]. p. 592-593.

cause already existing technologies and also potential future products will not have their place in the area of EU.

6. CONCLUSION

While analysing bans the goal is clearly visible, though vague and superficial regarding content. EU used the easier way to approach risks by banning AI systems instead of regulating how to use them or better said, regulating the intention of a creator behind the system. The same AI systems could be used to do harm or to benefit its users. The creator is the one who sets the parameters of the system which directs further actions. EU is fiercely focused on protecting human rights but putting stress on the human rights when not reasonable in such quantity can directly affect the competitiveness of the EU. During the research, it was noticed that lack of knowledge, when it comes to technologies, spins the question of human rights violations in a circle due to the fear of the unknown which may lead to huge loss in the area of innovations. If the EU wants to be competitive in the area of AI, it needs to find a better solution than a ban, regardless of the potential risk. Innovations are based on risk and to gain the most out of them we have to accept it. We can only imagine what the future will bring. Bans in the Proposal are definitely not the only potential risk which means that future possibilities are endless but that same risk arises from the question of how we use the technology we have. There is always someone whose intentions are not good but we as a society are so focused on the bad impact that we forget to look on the bright side and the progress we have made. All in all, the EU has recognized the risk but failed to present to us what particular technology is the one that can bring the described harm. Technologies have been described but in general terms and too abstract. Basically, technology may be invented but if the EU detects the slightest potential of harm the technology is banned.

7. BIBLIOGRAPHY

[1] Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence act) and amending certain Union legislative acts [online]. 2021. [cit. 02. 12. 2021]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[2] MILIŠA, Zlatko and NIKOLIĆ, Gabrijela. Subliminalne poruke i tehnike u medijima. *Nova prisutnost: časopis za intelektualna i duhovna pitanja*. Kršćanski akademski krug (KRAK), [online]. 2013, vol. XI, issue 2, p. 293-312. [cit. 02. 12. 2021]. ISSN 1334-2312, 1848-8676. Available at: https://hrcak.srce.hr/106397

[3] Act on Electronic Media; NN 111/21; Art. 21 (3), [online]. [b.r.] [cit. 02. 12. 2021]. Available at: https://www.zakon.hr/z/196/Zakon-o-elektroni%C4%8Dkimmedijima

[4] VEALE, Michael and BORGESIUS, Frederik Zuiderveen. Demystifying the Draft EU Artificial Intelligence Act.[online]. c 2021, p. 97-112 [cit. 02. 12. 2021]. Available at: https://arxiv.org/ftp/arxiv/papers/2107/2107.03721.pdf?fbclid=IwAR3q2rvj8xAw-pvTqZ5KVL97CCFidpfZXAI0xImNMXbbLdRBbAdWKeDGl6U

[5] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), [online]. 2005 [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/dir/2005/29/oj/eng

[6] PROPP, Kenneth and MACCARTHY, Mark. *Machines learn that Brussels writes the rules: The EU's new AI regulation* [online]. 2021. [cit. 02. 12. 2021]. Available at: https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/

[7] LUGURI, Jamie and STRAHILEVITZ, Lior. *Shining a Light on Dark Patterns*. Rochester, NY: Social Science Research Network, [online]. 2021. [cit. 02. 12. 2021]. DOI: 10.2139/ssrn.3431205

[8] Norwegian Consumer Council; *DECEIVED BY DESIGN- How tech companies use dark patterns to discourage us from exercising our rights to privacy*. [online]. 2018, p. 1-43 [cit. 02. 12. 2021]. Available at: https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-0627-deceived-by-design-final.pdf

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [online]. 2016. [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng

[10] RUCH, Simon, ZÜST, Marc Alain and HENKE, Katharina. Subliminal messages exert long-term effects on decision-making. *Neuroscience of Consciousness* [online]. 2016, vol. 2016, issue 1, p. 1-9 [cit. 02. 12. 2021]. ISSN 2057-2107. DOI: 10.1093/nc/niw013

[11] DEZFOULI, Amir, NOCK, Richard and DAYAN, Peter. Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences,* [online]. 2020, vol. 117, issue 46, p. 29221-29228 [cit. 02. 12. 2021]. ISSN 0027-8424, 1091-6490. DOI: 10.1073/pnas.2016921117

[12] Kapersky daily, Social scoring systems: current state and potential future implications. In: *Kasperksy daily* [online]. c 2021 [cit. 02. 12. 2021]. Available at: https://www.kaspersky.-com/blog/social-scoring-systems/

[13] NAST, Condé. The complicated truth about China's social credit system. *Wired UK* [online]. 2019. [cit. 02. 12. 2021]. ISSN 1357-0978. Available at: https://www.wired.co.uk/article/china-social-credit-system-explained

[14] CANALES, Katie. China's „social credit" system ranks citizens and punishes them with throttled internet speeds and flight bans if the Communist Party deems them untrustworthy. In: *Business Insider* [online]. 2021. [cit. 02. 12. 2021]. Available at: https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4

[15] ABC Finance. Surviving The Social Credit Score. In: *ABC Finance* [online]. 3. 12. 2021. [cit. 02. 12. 2021]. Available at: https://abcfinance.co.uk/blog/surviving-the-social-credit-score/

[16] EBERS, Martin et al. The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J*. Multidisciplinary Digital Publishing Institute, [online]. 2021, vol. 4, issue 4, p. 589-603 [cit. 02. 12. 2021]. ISSN 2571-8800. DOI: 10.3390/j4040043

[17] Council Framework Decision of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States (2002/584/JHA) [online]. 2009. [cit. 02. 12. 2021]. Available at: http://data.europa.eu/eli/dec_framw/2002/584/2009-03-28/eng

[18] Act on Protection of the Population from Infectious Diseases, NN 79/07 , 113/08, 43/09, 130/17 , 114/18 , 47/20 , 134/20, [online]. [b.r.]. [cit. 02. 12. 2021]. Available at: https://www.zakon.hr/z/1067/Zakon-o-za%C5%A1titi-pu%C4%8Danstva-odzaraznih-bolesti

[19] Criminal Law Act, NN 125/11, 144/12, 56/15, 61/15, 101/17, 118/18, 126/19, 84/21, [online]. [b.r.] [cit. 02. 12. 2021]. Available at: https://www.zakon.hr/z/98/Kazneni-zakon

[20] Innovatrics. Self-Isolation Rather than Quarantine - Thanks to Face Recognition. In: *Innovatrics* [online]. 18. 6. 2020. [cit. 02. 12. 2021]. Available at: https://www.innovatrics.-com/news/covid19-self-isolation-rather-than-quarantine-thanks-to-face-recognition/

[21] SIROTNIKOVA, Miroslava German. *Question Marks over Slovak Quarantine App Fuel Privacy Concerns* [online]. 2020. [cit. 02. 12. 2021]. Available at: https://balkaninsight.com/2020/05/20/question-marks-over-slovak-quarantine-app-fuel-privacy-concerns/

CHANGES IN UK-EU PERSONAL DATA TRANSFERS AFTER BREXIT[1]

ANNA TSUVINA[2]

> *The UK was a longstanding proponent of high data protection standards while part of the EU, and it will remain so as an independent nation, leading the way in creating the best possible data protection regime that exists globally.*
>
> DCMS, "*Data: A new direction*"[3]

1. INTRODUCTION

Since the United Kingdom (UK) left the European Union (EU) in 2020, the process of conducting personal data transfers has changed significantly. In particular, the UK is regarded as a third country in the context of Article 25 (1) of the General Data Protection Regulation (EU GDPR). UK-EU transfers, which are now regarded as cross-border personal data transfers, may be conducted only if the UK ensures an adequate level of data protection, namely, a level of protection of fundamental rights and freedoms that is essentially equivalent to that guaranteed within the EU.[4] The adequacy decisions were adopted by the EU Commission to settle the matter and make

[1] Esej byla zpracována v semestru podzim 2021 v rámci předmětu MVV1368K Privacy and-Personal Data na téma UK-EU Personal Data Transfers: Past, Present and the Future/ The essay was written in the autumn 2021 semester for the course MVV1368K Privacy and Personal Data on the topic of UK-EU Personal Data Transfers: Past, Present and the Future.

[2] Anna Tsuvina je studentkou na Yaroslav Mudryi National Law University, Faculty of Justice, kontakt: tsuvinaanna22@gmail.com

[3] The Government of the United Kingdom. Department for Digital, Culture, Media and Sport (DCMS). Public consultation on reforms to the UK's data protection regime. *Data: a new direction*, [online]. 10. 09. 2021, [cit. 05. 12. 2021]. p. 8, 123-124.

[4] *Art. 6 GDPR – Lawfulness of processing* [online] 2016. [cit. 05.12.2021]. Available at: https://gdpr-info.eu/art-6-gdpr/

the transfers possible and simplified after Brexit. At the same time, two important questions may arise. What is the role of these adequacy decisions? What are the future predictions for personal data transfers between the UK and the EU? This essay is devoted to identifying the past, present and future state of UK-EU personal data transfers. The attention is mainly focused on the UK adequacy decisions and their effect on the future of data transfers.

2. PAST

The history of UK-EU personal data transfers should be analyzed in the first place to demonstrate the change in the regulatory regime. To begin with, the UK was a part of the EU for almost fifty years, from 1973 to 2020. In this timeline, the problem of trans-border personal data transfers did not arise for the state as it was one of the Member States of the EU and all the transfers fell under the requirements of regulations that were in force for all the Member States. Specifically, the free flow of data was possible under Article 1 (2) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.[35] In addition, the EU GDPR, which also includes the provisions on the free flow of data within the EU, applied in the UK for almost two years, from 25 May 2018 to 31 January 2020. With the goal of implementing the EU GDPR, the UK adopted the Data Protection Act (DPA 2018), which is still one of the main regulations governing the usage of personal data and the flow of information in the state.[6]The DPA 2018 originally referred to the EU GDPR's most important provisions for the protection of personal data and adopted such main definitions used in the EU GDPR as “personal data”, “processing”, “data subject”, “controller”, “processor” etc. Therefore, the

[5] Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. [online] 24.10.1995 [cit. 05.12.2021]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046.

[6] Data Protection Act. [online] 2018 [cit. 05.12.2021]. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.

original provisions of the DPA 2018 demonstrated the clear intention of the national legislator to implement the EU GDPR into the domestic law of the UK.

The next period in the history of UK-EU personal data transfers was connected with the process of separation of the UK from the EU. On 23 June 2016, the UK held a referendum on its membership in the EU. The historic decision to leave the EU was reached in that referendum. On 31 January 2020 at midnight, when the Withdrawal Agreement entered into force, the UK left the EU.[7] In the context of data protection, the separation led to the situation where the EU GDPR, the main data privacy regulation throughout the EU, could no longer be applied in the UK. Instead, the UK GDPR was adopted to regulate the questions of personal data protection in the UK.[8] The DPA 2018 was amended to be read in conjunction with the new UK GDPR instead of the EU GDPR. Although mentioned regulations have much in common, there is one important distinguishing feature of the UK data protection framework. In particular, according to the UK GDPR and the DPA 2018, the Information Commissioner is the leading supervisor, regulator and enforcer of the UK GDPR.[9] The latest suggestions of the UK Government, which concern the Information Commissioner Office's (ICO) restructuring, deserve special attention in that regard. The government proposed to establish an independent board and a chief executive officer at the ICO. The board would be led by a chair with non-executive directors, while the chief executive officer would have responsibility for the running of the organization. Structural improvements were introduced to make the work of the supervisory authority more effective in the long term.

[7] Brexit: EU-UK relationship. In: EUR-Lex [cit. 05. 12. 2021]. Available at: https://eur-lex.europa.eu/content/news/Brexit-UK-withdrawal-from-the-eu.html

[8] Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) as it forms part of the law of England and Wales, Scotland and Northern Ireland by virtue of section 3 of the European Union (Withdrawal) Act 2018. [online]. 27. 4. 2016 [cit. 05. 12. 2021]. Available at: https://www.legislation.gov.uk/eur/2016/679/contents.

[9] The Government of the United Kingdom. Department for Digital, Culture, Media and Sport. Data Protection Act 2018 Factsheet – The Information Commissioner and Enforcement, [online]. 2018, [cit. 05. 12. 2021], p. 1.

On 1 January 2021, the EU-UK Trade and Cooperation Agreement (TCA) came into force, according to Article 201 (1) of which the EU and the UK were committed to ensuring cross-border data flows to facilitate trade in the digital economy.[10] In addition, in Article 525 (1) of the TCA was once again mentioned that onward transfers to a third country are allowed only subject to conditions and safeguards appropriate to the transfer ensuring that the level of protection is not undermined. Under the TCA, the EU and the UK also agreed on the interim solution (a bridging mechanism) to ensure the provisional continuation of personal data flow from the EU to the UK. In general, The TCA may be seen as the first step in the regulation of cross-border personal data transfers which was taken before the UK adequacy decisions were adopted in June 2021. The inclusion of the provisions on cross-border data flows helped to cut the loss of profits in the business sector and postpone the question for several months.

3. PRESENT

The current state of UK-EU personal data transfers is connected with the decisions of the EU Commission on the UK's adequacy under the EU GDPR and Law Enforcement Directive (LED).[11] In both decisions, the EU Commission stated that the UK ensures an adequate level of protection in the context of Article 25 (1) of the EU GDPR. This means that most data can continue to flow from the EU without the need for additional safeguards. At the same time, the so-called "sunset clause", which means that the UK adequacy decisions are limited to four years and will not be automatically renewed, was developed by the EU Commission. The new adequacy process will be required to determine whether the UK still ensures the essentially

[10] Trade and Cooperation Agreement between the European Union and the European Atomic Energy Community, of the one part, and the United Kingdom of Great Britain and Northern Ireland, of the other part [online]. 2020 [cit. 05. 12. 2021]. Available at: http://data.europa.eu/eli/agree_internation/2021/689(1)/oj/eng

[11] Commission Implementing Decision pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by the United Kingdom. [online] 28.06.2021 [cit. 05. 12. 2021]. Available at: https://ec.europa.eu/info/files/decision-adequate-protection-personal-data-united-kingdom-general-data-protection-regulation_en.

equivalent level of data protection in June 2025. In addition, during the four-year period, the EU Commission can amend, suspend, or repeal the adopted decisions if issues related to data protection that call into question the level of protection arise. There is also a possibility for the Court of Justice of the European Union to decide on the data protection level in the case an EU data subject or an EU data protection authority challenges these decisions.

In fact, although the value of positive adequacy decisions in allowing personal data to be transferred without any additional safeguards between the UK and the EU cannot be denied, they are just one of the mechanisms to enable such cross-border data transfers. To support trusted data flows across the world such alternative mechanisms as Standard Contractual Clauses (SCCs) are readily available, flexible and straightforward to implement.[12] However, a recent study estimated the costs of the absence of the UK adequacy decisions at around GB £1-1.6 billion (€1.116-1.7856 billion) for UK firms, stemming largely from companies reverting to alternative transfer mechanisms under the EU GDPR.[13] Therefore, the adequacy decisions may be considered in practice as one of the most effective tools to regulate cross-border data transfers compared to other alternatives. This explains the desire of the UK national authorities to get a positive adequacy decision despite all the doubts concerning the UK's relevant legislation, including those concerning public security, defence, national security, criminal law and the access of public authorities to personal data.

Specifically, according to some studies, UK surveillance activities do not fully comply with EU data protection and privacy standards. For instance, the UK Government Communications Headquarters (GCHQ) intercepts, retains and analyses masses of personal data by collaborating with or compelling private actors to provide access points. As Hendrik Mildebrath mentioned in the recent in-depth analysis for the European Parliamentary

[12] UK Business Data Survey 2021. In: *GOV.UK* [online] [cit. 05. 12. 2021]. Available at: https://www.gov.uk/government/statistics/uk-business-data-survey-2021

[13] European Parliament. Directorate General for parliamentary research services. *EU-UK private-sector data flows after Brexit: settling on adequacy : in depth analysis.* [online]. LU: Publications Office, 2020. [cit. 05. 12.2 021] p. 1, 15,17.

Research Service, the algorithmic detection used in the UK causes three main problems, namely the mathematically unavoidable fact of a large number of false positives or false negatives when searching for rare instances in large data sets ("base-rate fallacy"), built-in biases and opaque processing ("black box phenomenon"). In addition, the Investigatory Powers Act does not require the Investigatory Powers Commissioner to disclose intrusive data processing to the data subject, even where it would not jeopardize intelligence activities. So, these examples demonstrate the drawbacks in the regulation which confirm that the level of data protection in the UK may be seen as not essentially equivalent to that within the EU. Nevertheless, these particularities did not preclude the adoption of the adequacy decisions for the UK which include, inter alia, some rules on the usage of personal data by public authorities, notably for national security reasons. Furthermore, the adequacy decisions seem to be adopted on the basis of trustworthy relationships between the UK and the EU, taking into account their common historical background. As it was said in one of the recent official documents of the UK government, new arrangements to govern the continued free flow of personal data between the EU and the UK were needed as "part of the new, deep and special partnership".[14]

4. FUTURE

In the context of the future of UK-EU data flows several ideas should be highlighted. Firstly, the adequacy decisions seem to be an interim arrangement designed to make cross-border data transfers possible in the short term. As was already mentioned, they may be amended, suspended, and repealed. Secondly, the new adequacy decisions are highly questionable. It is still possible that the EU Commission will not adopt a new adequacy decision unless already mentioned issues of national security and surveillance regime will not be addressed by the government. Another challenge in this context is the intention of the UK government to allow free cross-border

[14] The Government of the United Kingdom. The exchange and protection of personal data: a future partnership paper. [online] [b.r.] [cit. 05. 12. 2021] p. 2.

data transfers with other states all over the world. Such a decision of the UK government may cause harm to the EU data protection system as the majority of mentioned states do not have the adequacy decisions. This may be seen as a gap in the closed system which is constructed within the countries that have the adequacy decisions and aims at the highest possible level of data protection among these third countries.

5. CONCLUSION

So, the history of UK-EU data transfers demonstrates that for a long time the regulatory regime stayed unchanged. As a Member State of the EU, the UK could count on the provisions for the free flow of data within the EU. After the separation from the EU, the TCA was adopted to make the transfers possible before the adoption of the adequacy decisions. Although the adequacy decisions were finally adopted by the EU Commission, the fact that some issues in the UK data protection framework are still visible today may not be neglected. This leads to uncertainty with regard to both already adopted and future adequacy decisions. However, the government still has four years to find the solution to the problem and improve the national strategy on how to keep the level of data protection in the state at the necessary level, namely, at the level that is essentially equivalent to that guaranteed within the EU.

6. BIBLIOGRAPHY

[1] The Government of the United Kingdom. Department for Digital, Culture, Media and Sport (DCMS). Public consultation on reforms to the UK's data protection regime. *Data: a new direction,* [online]. 10. 09. 2021, p. 1-146. [cit. 05. 12. 2021]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Refor m_Consultation_Document Accessible_.pdf.

[2] *Art. 6 GDPR – Lawfulness of processing* [online] 2016. [cit. 05. 12. 2021]. Available at: https://gdpr-info.eu/art-6-gdpr/

[3] Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. [online] 24. 10. 1995 [cit. 05. 12. 2021]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046.

[4] Data Protection Act. [online] 2018 [cit. 05. 12. 2021]. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.

[5] Brexit: EU-UK relationship. In: EUR-Lex [cit. 05.12.2021]. Available at: https://eur-lex.europa.eu/content/news/Brexit-UK-withdrawal-from-the-eu.html

[6] Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) as it forms part of the law of England and Wales, Scotland and Northern Ireland by virtue of section 3 of the European Union (Withdrawal) Act 2018. [online]. 27.4.2016 [cit. 05. 12. 2021]. Available at: https://www.legislation.gov.uk/eur/2016/679/contents.

[7] The Government of the United Kingdom. Department for Digital, Culture, Media and Sport. Data Protection Act 2018 Factsheet – The Information Commissioner and Enforcement, [online]. 2018, p. 1-4. [cit. 05. 12. 2021]. Available at: https://assets.publishing.service.gov-.uk/government/uploads/system/uploads/attachment_data/file/711238/2018-05-23_Factsheet_5_-_Information_Commissioner.pdf.

[8] Trade and Cooperation Agreement between the European Union and the European Atomic Energy Community, of the one part, and the United Kingdom of Great Britain and Northern Ireland, of the other part [online]. 2020 [cit. 05. 12. 2021]. Available at: http://data.europa.eu/eli/agree_internation/2021/689(1)/oj/eng

[9] Commission Implementing Decision pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by the United Kingdom. [online] 28.06.2021 [cit. 05. 12. 2021]. Available at: https://ec.europa.eu/info/files/decision-adequate-protection-personal-data-united-kingdom-general- data-protection-regulation_en.

[10] UK Business Data Survey 2021. In: *GOV.UK* [online] [cit. 05. 12. 2021]. Available at: https://www.gov.uk/government/statistics/uk-business-data-survey-2021

[11] European Parliament. Directorate General for parliamentary research services. *EU-UK private-sector data flows after Brexit: settling on adequacy : in depth analysis.* [online]. LU: Publications Office, 2021. p. 1-39 [cit. 05. 12. 2021]. Available at: https://data.europa.eu/doi/10.2861/595569

[12] The Government of the United Kingdom. The exchange and protection of personal data: a future partnership paper. [online] [b.r.] p. 1-15. [cit. 05. 12. 2021]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/639853/The_exchang e_and_protection_of_personal_data.pdf.