Kolarević Emina[*]
https://orcid.org/0000-0002-3270-6740

UDK: 004.8:342.727
Review article
DOI: 10.5937/ptp2201111K
Received: 25.02.2022.
Approved on: 08.03.2022
Pages: 111–126

THE INFLUENCE OF ARTIFICIAL INTELLIGENCE ON THE RIGHT TO FREEDOM OF EXPRESSION

ABSTRACT: The right to freedom of thought and expression represents one of the fundamental principles of a democratic and civilized society. The Internet has become the most important communication medium through which the individuals exercise their right to seek, receive and impart information and ideas of any kind, regardless of any frontiers. Various technologies have been used to enable an online communication, while today artificial intelligence systems are deployed in every corner of the Internet, providing information dissemination and communication. The application of the artificial intelligence systems is based on generating, collecting, and processing a large quantity of personal data with the aim of profiling users and predicting their future behaviour. This can have serious consequences for the right to freedom of expression. Through the content personalization on online platforms, particularly on social networks and search engines, the artificial intelligence systems choose the content that users can see and the order in which they see it, leaving them in the so-called 'filter bubbles'. Artificial intelligence systems also moderate the content, removing the one which does not comply with the rules of the online platforms, and, temporarily or permanently, blocking the users who violate the community rules, raising thus the issues of legality, legitimacy and proportionality of the decisions made by artificial intelligence.

[*] Teaching Associate, The Faculty of Law for Commerce and Judiciary, The University of Business Academy, Novi Sad, Serbia, PhD candidate, The Faculty of law, The University of Belgrade, Serbia, e-mail: emina.kolarevic@gmail.com

Key words: *Artificial intelligence, freedom of expression, freedom of thought, content moderation, content personalization.*

1. Introductory considerations

Artificial intelligence (hereinafter – AI), is dramatically changing every aspect of our lives. Its capabilities today include learning from experience and rendering decisions autonomously without human interference, which makes it a disruptive and most transformative technology of the early 21st century (Barfield, 2018, p. 2). It is present on our smart phones, in the form of voice assistants like Siri and Bixby, they can translate, read our e-mail, or search for a desired location. AI shapes our Internet experience every time when we are online, just like it selects movies and TV series for us on streaming services like HBO or Netflix, and we are one step away from producing the autonomous vehicles. Automated decision-making is increasingly used in the assessment of creditworthiness, in the selection of candidates in higher education, to assess the potentials of workers in the processes of employment and dismissals, and in diagnosing certain diseases.

Artificial intelligence systems rely on Big data, huge collections of data, collected from various and numerous sources, from interactions in social networks to data from street cameras. Estimates suggest that at least 2.5 quintillion bytes of data are produced worldwide each day (Price, n.d.), which requires a kind of sorting by AI and Big data technology allows for the widespread collection, storing and processing of data. Consequently, implications for the right to privacy and data protection are immense. Every online interaction creates a large number of data, each search on the Internet is recorded, including personal and sensitive data and collection of data mostly takes place without any knowledge and consent of the user. As a consequence, the possibility of abuse is great (Kostić, 2021, p. 18).

In addition to the obvious effect that collection and processing of data has on the right to privacy and data protection, the right to freedom of expression, whose exercise depends to a large extent on the privacy protection, may also be denied. “The technological development of modern society has led to the multiplication of the space in which social life and social interactions take place” (Bjelajac & Filipović, 2021, p. 16). The Internet is used every day to exercise freedom of expression, while social media ensure that everybody gets an opportunity to express his/her opinion; and seek, receive and impart information regardless of frontiers. However, AI curates and personalizes content available to users on search engines and

social media, based on data and personal preferences it has about users, just like it is responsible for content moderation, i.e. the removal of insulting and harmful content, as well as the content that does not comply with the rules of the online platforms.

This paper aims to examine the impact of the use of AI on the enjoyment of the right to freedom of opinion and expression. First, the definition of artificial intelligence will be presented, and the concept of algorithmic decision-making and the most advanced form of AI, machine learning, will be analysed, in order to better understand the ways in which they shape our reality. In the subsequent chapters, two functions of AI, personalization and moderation of online content will be explored, which are deployed by Internet intermediaries, particularly search engines and social networks, as well as their effects on the right to freedom of opinion and freedom to seek, receive and impart information and ideas. In the end, it is worth noting that the terms "AI" and "AI system" will be used in this paper, without specifying if they refer to algorithmic decision-making or machine learning.

2. Concept and forms of artificial intelligence

Artificial intelligence is used in various ways, fields of its application increase every day, while there is no consensus about its definition. John McCarthy, an American scientist and one of the founders of the AI discipline, defined AI in a conference in 1956 at *Dartmouth College* as "the science and engineering of making intelligent machines" (Liao, 2020, p. 3). In its broadest sense, it can be defined as making machines capable of performing tasks that require cognitive functions like thinking, learning and problem solving (ibid.). A detailed definition given by the EU High Level Expert Group on Artificial Intelligence (AI HLEG, 2018) reads: "Artificial intelligence refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions" (p. 7).

AI is thus an umbrella term describing the processes that, in the essence, partly or fully delegate activities of making and implementing decisions from people to software systems. Software systems act in the virtual world, and include search engines software, image analysis, virtual assistants, voice or

face recognition systems, while AI can be installed in hardware systems like autonomous cars, robots,[1] or drones (AI HLEG, 2018, p. 1).

AI includes a wide range of research and it can take various forms, the most advanced being machine learning, a technology that makes it possible for the systems to learn directly from examples, data and experience. Instead of following pre-programmed rules, as was the case with the earliest forms of AI, these systems "are set a task and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output" (The Royal Society, 2017, p. 19). For example, smart home devices or virtual assistants keep learning from collected data about everyday linguistic and speech patterns in order to process and answer more accurately requests from their users (Kaye, 2018, par. 4). Among many limitations of machine learning systems, very significant one lies in the fact that these systems are applied in the situations where moral decisions must be made without human supervision, although they do not understand characteristics of the real world: "They do not have self-awareness or consciousness and they cannot think for themselves" (Liao, 2020, p. 9).

AI systems are based on algorithms - sets of instructions with "encoded procedures for transforming input data into a desired outcome, based on specific calculations" (Gillespie, 2014, p. 167). They can be very simple, like those that arrange word lists in alphabetical order, or complex, as algorithms for voice recognition or financial forecasting. It is certainly important to underline that, although not all algorithms include AI, every AI system includes sets of algorithms. AI itself can be perceived as one complex algorithm that combines algorithms for performing various functions (Sartor & Lagioia, 2020, p. 3). Sometimes algorithm decides autonomously, like automatically filtering spam mail from the inbox, and sometimes people make decisions with the help of algorithm, for example, when hiring, and such decisions are semi-automated (Zuiderveen Borgesius, 2018, p. 11). It is precisely the algorithm decision-making that has the most significant implications for human rights, as the database that train the system can be biased and thus produces discriminatory effects. An example is Google's 'Photos', image recognition application, which labelled two dark-skin persons as "gorillas" in 2015, because the data used for training the algorithm did not include sufficient number of pictures of persons from various racial and ethnic groups (Simonite, 2018).

[1] For example, a humanoid robot of Hanson Robotics-a - Sophia, able to participate in interviews.

3. Content personalization, the right to freedom of opinion and the right to receive information and ideas

Today, when communication mostly takes place online, traditional media are being replaced by social networks, news aggregators and search engines, which dominate the ways in which individuals are informed. Society heavily relies on the online platforms to exercise freedom of expression and to access news, as the European Court of Human Rights confirms that Internet "plays an important role in enhancing the public's access to news and facilitating the dissemination of information in general" (Times Newspapers Ltd v. The United Kingdom, par. 27),[2] and "has now become one of the principal means by which individuals exercise their right to freedom of expression" (Ahmet Yıldırım v. Turkey, par. 54),[3] while "the same rights that people have offline must also be protected online, in particular freedom of expression" (UN Human Rights Council, 2021, par. 1). In the abundance of news and information available online, content personalization done by AI, organizes the chaos on the Internet, ensuring individuals find timely and relevant information (Kaye, 2018, p. 12).

Content personalization on online platforms implies the use of algorithms that filter news and information for users based on their personal preferences and wishes. Internet intermediaries, namely, carefully record and store data on each user's activity online, using these huge traces of data to create individual user models, on the basis of which the AI systems will predict what news and information is relevant for the user, filtering out and omitting irrelevant information. As a consequence, search engine results will vary from user to user, while individuals with the same friends on social networks will have access to different news and information based on their previous interaction with the platforms (Bozdag, 2013, pp. 209–211).

Internet search is one of the most widespread forms of content personalization done by AI (Kaye, 2018, p. 11). For this reason, search engines play a very important role for individuals seeking, receiving or imparting information (Committee of Experts on Internet Intermediaries, MSI-NET, 2018, p. 17). Google, for example, uses various 'signals' to personalize content, such as earlier searches and keywords, location, recent social network contacts, so "different users receive different results based

[2] Times Newspapers Ltd v. The United Kingdom, Applications nos. 3002/03 and 23676/03, Judgment of the ECtHR, 10 March 2009.

[3] Ahmet Yıldırım v. Turkey, Application no. 3111/10, Judgment of the ECtHR, 18 March 2013.

on the same keyword search" (Bozdag, 2013, p. 211). It is unlikely that the content that is not highly indexed or ranked by the search engine, will find its way to wider audience, or be seen at all (MSI-NET, 2018, p.17). On the other hand, social networks show the content based on subjective assessments of the extent to which it is attractive and interesting to the user (Kaye, 2018, p. 10). Facebook gets these assessments by registering and collecting data on the overall interaction history of users, i.e. user interaction with other users, the so-called 'social gestures', including likes, shares, comments, subscriptions to certain content, etc. (Bozdag, 2013, p. 211). Consequently, AI systems that perform personalization "thus control the incoming information (user does not see everything available), but also determine the outgoing information and who the user can reach (not everything shared by the user will be visible to others)" (Bozdag, 2013, p. 211).

The question arises as to how the personalized online experience, created by AI system based on personal preferences and interests, where users are offered little or no exposure to the opinions contrary to their own, jeopardizes individuals' right to form and develop their opinion. The freedom to hold opinion, as a separate right in the corpus of the rights that fall under the freedom of expression, is protected by all key international and regional human rights instruments, and was first proclaimed in the Universal Declaration of Human Rights (1948) in Article 19,[4] later reaffirmed by Article 19 of the International Covenant on Civil and Political Rights (1966) and the European Convention on Human Rights (1952) in Article 10. This freedom is necessary as a precondition for enjoying freedom of expression, since an individual will impart ideas and information that he previously had formed an opinion about (Beširević et al., 2017, p. 247). In addition, the freedom to hold opinion without interference is an absolute right, a *forum internum* of freedom of expression and as such "permits no exception or restriction" (UN Human Rights Committee, 2011, par. 9).

Furthermore, an essential element of the right to freedom of opinion is the right to form an opinion and develop it through judgment, as well as the "freedom from undue coercion in the development of an individual's beliefs, ideologies, reactions and positions" (Kaye, 2018, p. 23). It is also a duty of the state not to allow the indoctrination of its citizens, as well as the promotion of one-sided information that can constitute a serious obstacle to the freedom

[4] "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."

to form and hold opinions (Bychawska-Siniarska, 2017, p. 13). Traditional press influences the opinion of readers through content curation every day. Print media elevate stories for their cover pages, influencing knowledge of individuals about relevant daily events (Kaye, 2018, p. 24). However, the traditional press does not use "knowledge about each individual user to trigger fears and sensitivities related to certain topics with the aim of manipulating her" (Helberger, Eskens, van Drunen, Bastian & Moeller, 2020, p. 18), as is the case with personalized news.

Many scholars, namely, warn of the creation of 'filter bubbles' (Pariser, 2011) and 'echo chambers' (Sunstein, 2017) in which users can be isolated due to personalization of content. 'Filter bubble', a term coined by Eli Pariser (2011), denotes "a unique universe of information for each of us" (p. 9), in which recommender algorithms trap users in an invariable and rigid belief environment that matches their existing beliefs. (Nguyen, Hui, Harper, Terveen & Konstan, 2014, p. 678). This can result in reduced exposure to critical content and poor information on current social and political issues, as well as a missed opportunity to be faced with content from the opposite side of the political spectrum (Bozdag, 2013, p. 218). The creation of 'echo chambers', as a consequence of 'filter bubbles', restricts the individual to communicate only with like-minded persons, avoiding facts and beliefs he/she disagrees with (ibid.). In the Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (2019), the Council of Europe warned of the capability of the algorithmic decision-making and machine learning system to influence emotions and thoughts subconsciously, as well as the individual autonomy and the right to form opinions and make independent decisions (paras. 8–9). It is important to keep in mind that users cannot opt out of the AI personalization on online platforms, which can lead them to believe that the information and news that they get is the most relevant and objective information available (Kaye, 2018, p. 25).

Content personalization also raises the issue of the right to freedom to receive information. Freedom of expression is mostly associated with freedom of imparting information, while its essence is actually in receiving information: "Freedom to impart information is not related only to the freedom to send, but also takes into account the reader, listener, audience in general, and its right to freely receive information" (Beširović et al., 2017, p. 253). In a number of its judgments, European Court of Human rights emphasizes that the press has a duty to impart information and ideas about political and other issues of public interest, and that the public has the right to receive such information (The Sunday Times v. The United Kingdom, par. 65; Lingens v.

Austria, par. 41).[5] Article 10 of the ECHR, therefore, guarantees not only the freedom of press to inform the public, but also the right of the public to be duly informed (The Sunday Times v. The United Kingdom, par. 66),[6] while the freedom to receive information also implies that the right of an individual to receive information that others are willing to impart on him cannot be restricted (Leander v. Sweden, par 74).[7]

Thus, algorithmic decision-making in content personalization reduces diversity and quality of information that individuals receive, requested and received information is mostly a reflection and confirmation of existing opinions and attitudes, which leads to "never seeing the other side of an argument and thus fostering an ill-informed political discourse" (Bozdag, 2013, p. 218). In particularly sensitive societies, effects of 'filter bubbles' and 'echo chambers' can further exacerbate the socio-political climate and lead to further polarization and radicalization in the society (Kostić, 2021, p. 33).

4. Content moderation and the right to impart information and ideas

In September 2016 Facebook removed the cult Pulitzer Prize winning photograph "Terror of War", better known as "Napalm Girl", which features a girl running away naked after the napalm bomb attack during the Vietnam War. Facebook defended its decision by the fact that nudity violated community standards, particularly nudity in the context of child abuse images (Hu, Neupane, Echaiz, Sibal & Lam, 2019, p. 41). After a wave of negative publicity, Facebook annulled its decision and recognized the significance and value of the photo.

After certain content is uploaded in the digital space, it is automatically saved and available on various online platforms to a large number of users. The platforms, in the post-moderation process, through their own control mechanisms, assess whether the content is in compliance with the platform rules (*Terms of Service*). If non-compliance is determined, the content will be removed, while in case the conduct of a user violates community rules, the user will be temporarily or permanently blocked (Kostić, 2021, p. 35). It is precisely the AI systems that moderate content in accordance with the

[5] The Sunday Times v. The United Kingdom, Application no. 6538/74, Judgment of the ECtHR, 26 April 1979. Lingens v. Austria, Application no. 9815/82, Judgment of the ECtHR, 8 July 1986.

[6] The Sunday Times v. The United Kingdom, Application no. 6538/74, Judgment of the ECtHR, 26 April 1979.

[7] Leander v. Sweden, Application no. 9248/81, Judgment of the ECtHR, 26 March 1987.

standards and rules of social media platforms, removing and blocking content that threatens national security, incites violence, terrorism, constitutes hate speech, nudity, child exploitation, i.e. content which is prohibited under most laws. AI systems assess the content through upload filters and automatically block the publication of content that is prohibited according to predefined criteria, while there are also reporting mechanisms that allow users to report illegal and inappropriate content, followed by review procedures, which are decided by human moderators and/or AI, when the content can be removed, and user accounts temporarily or permanently blocked (Bukovska, 2020, pp. 32–33). The largest social media platforms, such as Google and Facebook, have often claimed that full removal of content is done with human intervention, while the truth actually is that moderation takes place through semi-automated or fully automated processes (MSI-NET, 2018, p. 18).

According to online platforms, AI systems are not only more efficient in identifying illegal content and content that does not comply with the rules of the platform, but they also have a higher accuracy rate than human decision-making (Kaye, 2018, p. 14). The reason for this lies in the amount of inappropriate, harmful, and illegal content that is dispatched to the online space every hour, which exceeds capacities of human moderation (Kaye, 2018, p. 14). It should also be taken into consideration that the moderation of harmful content implies encounters with very disturbing scenes, such as scenes of explicit violence, child abuse and other toxic imagery, which proved to be very challenging for humans, both psychologically and in terms of health (York & Zuckerman, 2019, p. 149).

On the other side, the limitations of AI in content moderation are not insignificant. First of all, AI does not have the capacity to assess the context and nature of content and take into account different variations of language cues, as well as linguistic and cultural specificities (Kaye, 2018, p. 15). In terms of development, it has not yet reached a level where it can "differentiate between news reporting, advocacy, and satire on the one hand, and on the other, the actual incitement of harm" (Hu, Neupane, Echaiz, Sibal & Lam, 2019, p. 38). That is why, when assessing the context of content, AI is prone to errors: it can identify illegal content as permissible, which results in 'false negatives', or 'false positives' when removing legitimate content (Bukovska, 2020, p. 56). For example, in 2017, videos that showed alleged war crimes in Syria, documented by journalists and activists, were identified by YouTube machine learning systems as terrorist propaganda and they were removed (Citron & Jurecic, 2018, p. 13). The use of AI in detecting content that can pose a threat to national security can be problematic indeed. First, there is no

unique definition of such content, while terms like 'terrorism' or 'extremism' are not uniformly defined under international human rights law (Bukovska, 2020, p. 55). Second, videos and other content may advocate terrorism in one context, while in another, they may be crucial for news reporting or combating the recruitment of terrorists online (Bukovska, 2020, p. 56).

It is similar with hate speech. International human rights instruments define and stipulate restrictions on hate speech in different ways: "the context of the speech, the role of the speaker and the likelihood that the speech results in harm are among the key factors to determine whether the speech in question should be restricted" (Bukovska, 2020, p. 58). Public debates about issues of public interest are often heated and accompanied by offensive or figurative speech, irony, and mockery, which does not constitute hate speech, but AI systems can easily identify it as such and remove it: "natural language processing cannot differentiate between a sarcastic rebuttal of hate speech from an actual hateful comment" (Hu, Neupane, Echaiz, Sibal & Lam, 2019, p. 40). AI is, therefore, blind to such contextual differences.

It follows from the above that the AI may excessively remove legitimate content, which raises the issue of interference into the enjoyment of freedom of expression, as well as the "questions of legality, legitimacy and proportionality" (MSI-NET, 2018, p. 18). Namely, freedom to impart information and ideas has always been considered a key part of freedom of expression. UDHR in Article 19 and ICCPR in Article 19 guarantee the freedom to impart information and ideas by any means and regardless of frontiers, while Article 10 of the ECHR underlines that every person enjoys this right without interference of public authorities. The exercise of freedom of expression depends not only on the obligation of public authorities not to interfere, but also demands "positive measures of protection, even in the sphere of relations between individuals" (Ozgur Gundem v. Turkey, par. 43).[8] Social media platforms are under pressure from states and international community, and are often obliged, depending on the jurisdiction they are located in, to remove illegal and harmful content, which can lead to excessive removal of content that is not only harmless, but significantly contributes to public debate (MSI-NET, 2018, p. 21). In that respect, it is important to keep in mind that freedom of expression protects not only "information or ideas that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands

[8] Ozgur Gundem v. Turkey, Application no. 23144/93, Judgment of the ECtHR, 16 March 2000.

of that pluralism, tolerance and broadmindedness without which there is no democratic society" (Handyside v. United Kingdom, par. 49).[9]

Furthermore, algorithmic content removal and account blocking also raises the issue of legitimacy of interference with the freedom of expression. According to Article 10 paragraph 2 of the ECHR, freedom of expression can be subject to restrictions when they are prescribed by law and necessary in a democratic society for achieving one of legitimate social goals. It follows that content removal and blocking constitutes a restriction on the freedom to impart information and ideas, which should correspond to a 'pressing social need' and pass the proportionality test, i.e. it should be proportionate to the legitimate goal. In the case of *Ahmet Yıldırım v. Turkey*[10] (2012), European Court of Human Rights established that the measure blocking Google sites in Turkey "produced arbitrary effects" (par. 68) and restricted the rights of users; had a significant collateral effect as it made large amounts of information inaccessible (par. 66), "for news is a perishable commodity and to delay its publication, even for a short period, may well deprive it of all its value and interest" (par. 47). In addition, removing online content or blocking Internet access require a legal framework that would provide strict control over the scope of restraints and effective judicial revision aimed at preventing abuse of power, while "the judicial review of such a measure, based on a weighing-up of the competing interests at stake and designed to strike a balance between them, is inconceivable without a framework establishing precise and specific rules regarding the application of preventive restrictions on freedom of expression" (par. 64).

Social media companies, however, are invited and encouraged to voluntarily remove any illegal content from their platforms without a clear legal basis, which further complicates the provision of basic legal guarantees, such as accountability and transparency (MSI-NET, 2018, p. 19). Inviting Internet intermediaries to set their own criteria for removal or blocking of content, can be dangerous in terms of determining and implementing solutions by the private sector, which public authorities themselves were not able to legally prescribe: "public-private partnerships may thus allow public actors to impose regulations on expression that could fail to pass constitutional muster in contravention of rule of law standards" (MSI-NET, 2018, p. 20).

[9] Handyside v. United Kingdom, Application no. 5493/72, Judgment of the ECtHR, 7 December 1976.

[10] Ahmet Yıldırım v. Turkey, Application no. 3111/10, Judgment of the ECtHR, 18 March 2012.

5. Conclusion

Freedom of expression achieves its full application on the Internet, on online platforms, which not only allow the communication of users and expression of their opinions, but also enable the access to abundance of information and knowledge, which is decisive for forming and developing opinion. Search engines, social networks, online media, and news aggregators have a very important role for individuals who seek, receive, and impart information and ideas.

However, regardless of the significant improvements that artificial intelligence brings to our society, its unfavourable impact on the freedom of opinion and expression should not be neglected. Through personalization and recommender systems, AI shapes the online experience and decides which information and news the users will have access to, on the basis of data collected about them in a previous interaction with the platform. Content personalization has certain advantages, like having the access to information that are specifically tailored to the needs of the user, what, on the other hand, violates the right of the individual to have access to all relevant news and information of general interest. This also raises the question of interfering with the right to hold an opinion and freedom from undue coercion in forming and developing one's own opinion, bearing in mind that the result of content personalization may be the creation of the filter bubbles and echo chambers, where users have a restricted scope of accessible information that mostly reflect their earlier beliefs and attitudes. On the other side, content moderation practices, which entail removing of harmful and illegal content from social media platforms that are prohibited under international human rights law and are not in line with the rules of online platforms, may also lead to excessive removal of perfectly legal content and speech due to the fact that AI cannot recognize speech context and make a difference between, for instance, actual hate speech and irony or satire. Algorithmic censorship of online content, carried out by Internet intermediaries due to public pressure and frequently without legal basis, raises thus the issues of legality, legitimacy, and proportionality of restricting the freedom to impart information. The lack of legal framework that establishes precise rules for restricting freedom of expression online is yet another very important challenge of content moderation.

Kolarević Emina
Pravni fakultet za privredu i pravosuđe, Univerzitet Privredna akademija u Novom Sadu, Srbija

UTICAJ VJEŠTAČKE INTELIGENCIJE NA UŽIVANJE PRAVA NA SLOBODU IZRAŽAVANJA

REZIME: Pravo na slobodu mišljenja i izražavanja predstavlja jedno od temeljnih načela demokratskog i civilizovanog društva. Internet je danas postao najznačajniji komunikacioni medij pomoću kojeg pojedinci ostvaruju svoje pravo da traže, primaju i saopštavaju informacije i ideje svih vrsta, bez obzira na granice. Razne tehnologije su korišćene da bi se omogućila onlajn komunikacija, dok se danas sistemi vještačke inteligencije nalaze u svakom kutku interneta, obezbjeđujući diseminaciju informacija i komunikaciju. Primjene sistema vještačke inteligencije zasnivaju se na generisanju, prikupljanju i obradi velike količine ličnih podataka u cilju profilisanja i predviđanja budućeg ponašanja korisnika, što može imati ozbiljne posljedice po pravo na slobodu izražavanja. Sistemi vještačke inteligencije putem personalizacije sadržaja na onlajn platformama, pogotovo na socijalnim mrežama i internet pretraživačima, biraju koji sadržaj korisnici mogu da vide i kojim redom ga vide, ostavljajući ih u takozvanim 'fitler mjehurima'. Sistemi vještačke inteligencije takođe vrše moderaciju sadržaja, uklanjajući sadržaj koji nije u skladu sa pravilima onlajn platformi i blokirajući privremeno ili trajno korisnike koji se ponašaju suprotno pravilima zajednice, što povlači pitanja zakonitosti, legitimnosti i proporcionalnosti odluka vještačke inteligencije.

Ključne riječi: *vještačka inteligencija, sloboda izražavanja, sloboda mišljenja, moderacija sadržaja, personalizacija sadržaja.*

References

1. Barfield, W. (2018). Towards a law of AI. In: Barfield W. & Pagallo U. (eds.), *Research Handbook on the Law of AI* (pp. 2–39). Cheltenham & Northampton: Edward Elgar Publishing

2. Beširević, V. et al. (2017). *Komentar Konvencije za zaštitu ljudskih prava i osnovnih sloboda* [*Commentary on the Convention for the Protection of Human Rights and Fundamental Freedoms*]. Beograd: JP Službeni glasnik
3. Bjelajac, Ž. & Filipović, A. (2021). Specific characteristics of digital violence and digital crime. *Pravo - teorija i praksa*, *38* (4), pp. 16–32, DOI: 10.5937/ptp2104016B
4. Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15* (3), pp. 209–227. DOI: 10.1007/s10676-013-9321-6
5. Bukovska, B. (2020). *Spotlight on Artificial Intelligence and Freedom of Expression*. Vienna: OSCE Representative on Freedom of the Media. Downloaded 2021, October 26 from https://www.osce.org/files/f/documents/9/f/456319_0.pdf
6. Bychawska-Siniarska, D. (2017). *Protecting the Right to Freedom of Expression under the European Convention on Human Rights*. Strasbourg: Council of Europe. Downloaded 2021, October 21 from https://rm.coe.int/handbook-freedom-of-expression-eng/1680732814
7. Citron, D. K. & Jurecic, Q. (2018). *Platform Justice: Content Moderation at an Inflection Point*. Hoover Institution. Downloaded 21, October 6 from https://www.hoover.org/sites/default/files/research/docs/citron-jurecic_webreadypdf.pdf
8. Committee of Experts on Internet Intermediaries [MSI-NET] (2018). *Algorithms and Human Rights Study on The Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications* (DGI(2017)12). Strasbourg: Council of Europe. Downloaded 2021, March 25 from https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5
9. Council of Europe (1952). *The European Convention on Human Rights.* Strasbourg: Directorate of Information. Downloaded 2021, October 16 from https://www.echr.coe.int/documents/convention_eng.pdf
10. Council of Europe (2019). *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*. Strasbourg. Downloaded 2021, November 2 from https://www.coe.int/en/web/data-protection/-/declaration-by-the-committee-of-ministers-on-the-manipulative-capabilities-of-algorithmic-processes
11. European Commission's High-Level Expert Group on Artificial Intelligence [AI HLEG] (2018). *A Definition of AI: Main Capabilities and Scientific Disciplines*. Brussels: European Union. Downloaded 2021,

October 22 from https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf
12. Gillespie, T. (2014). The Relevance of Algorithms. In: Gillespie T, Boczkowski, P. J. & Foot, K. A. (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167–193). Cambridge, Massachusetts: Massachusetts Institute of Technology
13. Helberger, N., Eskens, S., van Drunen, M., Bastian, M. & Moeller, J. (2020). *Implications of AI-driven Tools in the Media for Freedom of Expression*. Conference of Ministers responsible for Media & Information Society, Nicosia, Cyprus. Downloaded 2021, May 10 from https://rm.coe.int/cyprus-2020-ai-and-freedom-of-expression/168097fa82
14. Hu, X., Neupane, B., Echaiz, L. F., Sibal, P. & Lam, M. R. (2019). *Steering AI and Advanced ICTs for Knowledge Societies: A Rights, Openness, Access and Multi-stakeholder Perspective*. Paris: UNESCO Publishing. Downloaded 2021, March 4 from https://unesdoc.unesco.org/ark:/48223/pf0000372132
15. Kaye, D. (2018). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* (A/73/348). New York: United Nations General Assembly. Downloaded 2021, March 24 from https://ap.ohchr.org/documents/dpage_e.aspx?si=A/73/348
16. Kostić, B. (2021). *Vještačka inteligencija: uticaj na slobodu izražavanja, medijske perspektive i regulatorni trendovi* [*Artificial Intelligence: The influence on freedom of expression, media perspectives and regulatory trends*]. Belgrade: Organization for Security and Co-operation in Europe. Downloaded 2021, October 26 from https://www.osce.org/sr/mission-to-serbia/479672
17. Liao, S. M. (2020). A Short Introduction to the Ethics of AI. In: Liao S. M. (ed.), *Ethics of AI* (pp. 1–42). Oxford: Oxford University Press
18. Nguyen, T. T., Hui, P. M., Harper, F. M., Terveen, L. & Konstan, J. A. (2014). Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. *Proceedings of the 23rd International Conference on World Wide Web* (pp. 677–686). DOI: 10.1145/2566486.2568012
19. Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin Books
20. Price, D. (n.d.). *Infographic: How Much Data is Produced Every Day?* CloudTweaks. Downloaded 2021, October 21 from https://cloudtweaks.com/2015/03/how-much-data-is-produced-every-day/
21. The Royal Society (2017). *Machine Learning: the power and promise of computers that learn by example.* Downloaded 2021, October 21

from https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf
22. Sartor, G. & Lagioia, F. (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. Directorate-General for Parliamentary Research Services of the Secretariat of the European Parliament. Brussels: European Union. DOI: 10.2861/293
23. Simonite, T. (2018). When It Comes to Gorillas, Google Photos Remains Blind. *Wired*, November 1. Downloaded 2021, October 20 from https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/
24. Sunstein, C. R. (2017). *Republic: Divided Democracy in the Age of Social Media*. Princeton and Oxford: Princeton University Press
25. UN Human Rights Committee (2011). *General Comment No. 34 Article 19: Freedoms of opinion and expression*. Geneva, Switzerland. Downloaded 2021, November 1 from https://www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf
26. UN Human Rights Council (2021). *The promotion, protection and enjoyment of human rights on the Internet* (A/HRC/47/L.22). New York: United Nations General Assembly. Downloaded 2021, November 1 from https://undocs.org/A/HRC/47/L.22
27. United Nations (1966). *International Covenant on Civil and Political Rights. Treaty Series, 999*. Downloaded 2021, October 16 from https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx
28. United Nations (1948). *Universal Declaration of Human Rights. United Nations General Assembly*. Paris, France. Downloaded 2021, October 16 from https://www.un.org/en/about-us/universal-declaration-of-human-rights
29. York, J. C. & Zuckerman, E. (2019). Moderating the Public Sphere. In: Jørgensen R. F. (ed.), *Human Rights in the Age of Platforms* (pp. 137–161). Cambridge, Massachusetts: Massachusetts Institute of Technology
30. Zuiderveen Borgesius, F. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making*. Strasbourg: Council of Europe. Downloaded 2021, September 23 from https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-%20making/1680925d73