Respect for Personal Autonomy in AI Regulatory Framework[1]

Petro M. Sukhorolskyi*

Lviv Polytechnic National University,
Lviv, Ukraine
*e-mail: petro.m.sukhorolskyi@lpnu.ua

Abstract
The legal regulation of artificial intelligence is one of the most pressing and debated topics at the national and international levels. The rapid development of artificial intelligence can significantly change the existing reality and leads to fundamentally new challenges for lawmaking and law enforcement, in particular in the field of human rights. The main purpose of the article is to determine whether the new European legal instruments on artificial intelligence (in particular, the European Union's AI Act and the Council of Europe's Framework Convention on AI) reflect these technological threats and protect the personal autonomy of individuals. To achieve this goal, the article reveals the essence of personal autonomy and its significance for human rights and the legal system, as well as identifies the directions of the real and potential impact of artificial intelligence on personal autonomy. The theoretical and methodological foundation of the study is Joseph Raz's theory of personal autonomy which allows to identify the main problems and contradictions in the use of artificial intelligence and to shape proposals for responding to actual threats. Based on the idea of the fundamental role of personal autonomy, the article shows how the introduction of artificial intelligence, driven by the interests of specific actors, negatively affects the position, rights and capacities of individuals. In particular, the author identifies three directions of such influence: high-tech manipulation of people, distortion of their perception through myths and misconceptions, and formation of the appropriate online architecture and social norms. Based on the analysis of legal documents, two approaches to the regulation of artificial intelligence are identified. The first approach relegates personal autonomy to the periphery and suggests that problems should be solved through cooperation between government and business by using risk assessment tools. This should result in ready-made solutions that are offered to people.

The second human-centred approach emphasises the protection of personal autonomy. However, detailed norms within this approach have not yet been created, and their development requires further theoretical elaborations. In this regard, the primary focus should be on preserving and improving the conditions of autonomy that are threatened by the misuse of artificial intelligence.

Keywords: personal autonomy; artificial intelligence; human rights; personal data protection; negative and positive freedom.

Повага до персональної автономності в системі правового регулювання штучного інтелекту

Петро Михайлович Сухорольський*

*Національний університет «Львівська політехніка»,
Львів, Україна
e-mail: petro.m.sukhorolskyi@lpnu.ua

Анотація

Тема правового регулювання штучного інтелекту є однією з найактуальніших і найбільш обговорюваних на національному та міжнародному рівнях. Швидкий розвиток штучного інтелекту здатний суттєво змінити навколишню реальність та зумовлює появу принципово нових викликів для правотворчості й правореалізації, зокрема у сфері прав людини. Мета цієї статті полягає у встановленні того, наскільки нові європейські правові інструменти щодо регулювання штучного інтелекту, зокрема Регламент Європейського Союзу та Конвенція Ради Європи, відображають указані технологічні загрози та захищають персональну автономність індивідів. Для досягнення цієї мети розкрито сутність персональної автономності та її значення для прав людини і для правової системи загалом, а також встановлено напрями реального та потенційного впливу штучного інтелекту на персональну автономність. Теоретико-методологічну основу дослідження становить теорія персональної автономності Джозефа Раза, яка дає змогу виявити основні проблеми та суперечності у сфері використання штучного інтелекту, а також сформувати пропозиції щодо реагування на актуальні виклики. Опираючись на теорію про фундаментальну роль персональної автономності, у статті продемонстровано, як сучасні процеси впровадження штучного інтелекту, керовані інтересами конкретних суб'єктів, негативно впливають на становище, права та можливості індивідів. Виокремлено три напрями такого впливу: високотехнологічні маніпуляції над людиною, викривлення її сприйняття через міфи і необґрунтовані уявлення, формування відповідної онлайн-архітектури і соціальних норм. На основі проведеного аналізу правових актів визначено два підходи до регулювання штучного інтелекту. У межах першого з них персональна автономність зміщена на периферію і проблеми вирішують через співпрацю влади і бізнесу за допомогою процедур оцінки ризиків. Другий, людиноцентричний, підхід ставить на чільне місце

захист персональної автономності. Проте деталізовані норми в його межах ще не створені, а для їх розроблення потрібні напрацювання на теоретичному рівні. Впроваджуючи цей підхід, основну увагу потрібно надавати збереженню і покращенню умов автономності, які опиняються під загрозою внаслідок недосконалих практик використання штучного інтелекту.

Ключові слова: персональна автономність; штучний інтелект; права людини; захист персональних даних; негативна і позитивна свобода.

Introduction

The emergence of the Internet, the development of technology and the shift of many activities online have necessitated the transformation of legislation and the development of new legal instruments. Initially, there was a tendency to apply the same principles and norms that govern relations in the "real" world to online relations gradually adapting these norms to the Internet environment. However, nowadays, especially with the rise of artificial intelligence (hereinafter: AI), there is a growing understanding that more fundamental changes in terms of legal regulation are needed, including a reassessment of the legal principles and values underlying various legal institutions. Personal (individual) autonomy (hereinafter: PA) is one of these values which directly affects the possibility of exercising individual's rights. Despite its fundamental importance and the attention paid to it by a number of prominent legal philosophers over the centuries, it has not been adequately reflected in constitutional acts and international human rights documents. Many modern researchers believe that individual autonomy faces the greatest threat due to widespread digitalization, automation, and the introduction of intelligent algorithms that increasingly affect people and society.

In general, the topic of AI legal regulation is one of the most pressing and discussed both at the national and international levels. In 2024 alone, two fundamental legal documents have been adopted in Europe on this issue: The Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (hereinafter: AI Act) [1] and the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (hereinafter: Convention on AI) [2]. In the near future, relevant legislative acts will appear in many countries of the world, including Ukraine, that, in the course of its accession negotiations with the EU, must implement the acquis and is likely to become a party to the Council of Europe Convention. To develop and implement such acts, it is necessary to have a good understanding of this area, but two factors hinder this. Firstly, the attitude among experts towards the potential of AI and its prospects is not unambiguous. Secondly, this issue affects the interests of many influential

economic and political actors, which gives rise to significant information distortions and manipulations. This study focuses on one aspect of the problem, namely, the protection of the rights and interests of individuals in the context of AI regulation. At the heart of these rights and interests there lies the possibility of personal autonomy which is often forgotten when considering the impact of informatisation on human rights. The fundamental importance of personal autonomy has been emphasized by numerous human rights researchers, primarily Raz [3], Dworkin [4], Bernal [5] et al. Their ideas are rooted in the theories of freedom of the classics of liberal thought, in particular John Stuart Mill [6] and Isaiah Berlin [7]. Nowadays, the issue of AI's impact on human rights is being studied by Risse [8], Rodrigues [9], Shaelou, Razmetaeva [10] and others. Specifically, Prunkl [11] and Mik [12] have conducted detailed studies on the threat posed by modern information technologies and AI to human autonomy. However, this topic remains under-researched, especially given the fundamental importance of PA for the legal system. In addition, the aforementioned works do not provide an analysis of the AI regulations adopted in 2024 from the perspective of PA.

Given all of the above, the main objective of this article is to establish to what extent the new European AI regulations reflect current technological threats and protect individuals' personal autonomy in the context of the increasing penetration of these technologies into our reality. To do this, it is first necessary to clarify the essence of PA and its significance for human rights and the legal system in general, as well as to define how AI algorithms affect or can potentially affect PA.

Materials and Methods

The theoretical and methodological foundation of this study is primarily the theory of personal autonomy by the renowned legal philosopher Joseph Raz, as outlined in his key work "The morality of freedom" [3]. According to Raz, personal autonomy is a fundamental component and ultimate value within modern legal systems based on the principles of liberal democracy. Freedom is obviously a central value in a liberal state, but the existence of various, often contradictory, interpretations of it, especially those related to the issue of positive freedom, makes it very difficult to define specific rights and obligations and the scope of state intervention in this area. Raz notes that positive liberty "like all notions which have become slogans in intellectual battles, is notoriously elusive" [3, p. 409]. However, if we realize that it is based on the phenomenon of personal autonomy, then the situation becomes much clearer. That is why he is convinced that "positive freedom derives its value from its contribution to

personal autonomy" and "[d]isputes concerning the scope and content of positive freedom should be settled by reference to the contribution of the disputed element to autonomy" [3, p. 409].

The above also applies to determining the conditions under which the state may resort to coercion in accordance with the well-known harm principle (which implies that the actions of individuals may be limited only to prevent harm to other individuals). Accordingly, the state may use coercion (as well as manipulation), i.e. reduce the personal autonomy of one person, only when it is justified by the protection of the personal autonomy of other persons [3, p. 416]. In this respect, the protection of the PA of others means not only ensuring the absence of interference, but also creating conditions of autonomy. Such conditions are complex and multidimensional and include at least three components: 1) the existence of mental capacities that allow people to be autonomous, 2) an adequate range of available options, and 3) relative independence that excludes living under coercion or manipulation, except when they are necessary to ensure the personal autonomy of others [3, p. 373]. Thus, the state or any authority has no right to justify coercion or manipulation of people, i.e. treating them as objects rather than autonomous persons, with other goals, values and arguments than protection of PA.

Although Raz's theory arose long before the emergence of the modern commercialised Internet, where large-scale data collection and AI algorithms are implemented, it is best suited to identify the main problems and contradictions related to the AI regulation and to develop adequate proposals for responding to the existing challenges. Such an understanding of PA helps to establish how the current AI development and implementation processes, driven by the interests of influential actors in this area, negatively affect human autonomy. To prove this, the article uses general research methods, including analysis, synthesis, generalisation, and induction. The empirical basis of the study are the opinions of experts in the field of AI and its regulation, set out in reports, interviews, and media publications. Besides, the formal legal method and the comparative legal method are used to analyse and compare the provisions of international legal instruments on AI regulation.

Results and Discussion

The concept of personal autonomy and its importance for human rights and the legal system

The idea of the centrality of PA to the political and legal system was shared by many prominent liberal philosophers of the past centuries, such as Thomas Paine, John Stuart Mill, Isaiah Berlin. However, not all of them directly used the word

"autonomy" in their works to denote a certain quality that makes a person the master of his or her own life [5]. Mill's understanding of freedom is the closest to the PA. In particular, he notes that "[t]he only freedom which deserves the name, is that of pursuing our own good in our own way" and "Mankind are greater gainers by suffering each other to live as seems good to themselves, than by compelling each to live as seem good to the rest" [6, p. 83]. According to Mill, the principle of liberty "requires liberty of taste and pursuits; of framing the plan of our life to suit our own character; of doing as we like, subject to such consequences as may follow: without impediment from our fellow-creatures, so long as what we do does not harm them, even though they should think our conduct foolish, perverse, or wrong" [6, p. 83]. Paul Bernal, a professor of information technology law, believes that autonomy can be considered the most important aspect of Mill's political philosophy [5, p. 30].

The idea of autonomy also permeates Isaiah Berlin's landmark work "Two Concepts of Liberty", in which the philosopher distinguishes between negative and positive freedom, i.e. freedom from interference and freedom to do something in the external world. Among other things, Berlin notes that positive freedom implies splitting of "personality into two: the transcendent, dominant, controller, and the empirical bundle of desires and passions to be disciplined" [7, p. 111]. He believes that this has contributed to the blurring of the concept of freedom and has led to numerous manipulations, up to the point where man and freedom can mean "whatever the manipulator wishes" [7, p. 111].

This problem directly leads us to autonomy. The concept of PA implies that a person is the master of his own life, the sovereign of his own happiness, and he himself must make decisions about his own destiny and take responsibility for them [4, p. 112]. Autonomy presupposes that a person has many goals and interests at the same time, even those that conflict with each other. And he can make different choices at different times, including changing his preferences [4, p. 17]. With this in mind, it is clear how to resolve the conflict described by Berlin: by creating conditions of autonomy so that a person can make decisions for himself. In this context, Gerald Dworkin, in his book on autonomy, considers the example of a smoker who has both the desire to smoke and the desire not to have that desire [4, p. 15]. Which of these two conflicting desires is then to be considered the true expression of that person's will? And should the state or anyone else use coercion and manipulation to make people stop smoking, and thereby help them fulfil their deep desire? Based on the value of freedom, given its various interpretations, we cannot give unambiguous answers to these questions. But from the perspective of personal autonomy, everything seems much simpler. It is quite natural for a person to have multiple and conflicting

goals, and no one has the right to impose any decision on others by force. The government can only protect non-smokers from second-hand smoke and promote healthy lifestyles among everyone, leaving the final choice to the individual smoker.

If we compare this situation with the consumption of content on the Internet, we need to take into account that algorithms on websites make decisions for people based on detailed information about them. They usually pay attention to the superficial level of one's interests, ignoring deeper goals and aspirations. For example, YouTube tracks users' behaviour and pushes similar content to them. It has been recorded that people spend 70 % of their "view-time" watching videos recommended by the algorithm [13]. From YouTube's point of view, such a high rate is evidence of the perfection of the algorithm, but if we consider the value of PA, the question arises whether people really want to watch this content. It is quite reasonable to assume that without intrusive algorithms, they would search for and choose something else, just as they do in the real world, but the system constantly channels their choices in a specific direction, exploiting their weaknesses.

The right to be forgotten, a newly emerged right in Europe that has stirred up considerable debate in the expert community and created a watershed between European and American understandings of Internet freedom, is also aimed at protecting personal autonomy on the Internet [14]. This right is intended to protect a person from the pressure of a technological system that has accumulated a lot of information about people, and from the pressure of society which has access to some of this information through search engines. The case of Mario Costeja González, which led to the EU Court of Justice's decision establishing the right to be forgotten, shows that a person is judged by what is visible on the surface (often scandalous or negative information from the past) without taking into account that person's complex personality and the dynamics of its development [15]. In this way, the individual is firmly embedded in the existing information structures that deprive him of autonomy. Instead, the right to be forgotten is intended to restore this autonomy and allow a person to start life from scratch and freely assert himself in the external world in various manifestations and variations. The absence of a clear recognition of the right to autonomy in legal regulations made authors of the right to be forgotten rely primarily on the right to privacy, which is open to different interpretations. This is the main reason for the lack of consensus on the right to be forgotten. After all, although privacy, as well as freedom, intersect with autonomy to some extent, there are differences between them that can lead to a completely different outcome when balancing rights and interests.

According to a number of theorists, PA plays a central role in modern normative philosophy [4, p. 4]. Ratz argues that it is a mistake to consider autonomy as just one of many important values, since it constitutes "a central aspect of the whole system of values of a society, which affects its general character" [3, p. 395]. That is why it needs to be conceptualised in no less detail and be provided with no weaker legal basis than freedom or equality [4, p. X]. The core of this process is the creation and improvement of conditions of autonomy, which is the government's main task. The first component of these conditions is the mental capacity of a person to be autonomous [3, p. 369]. In other words, we need an environment that fosters independent thinking and the ability to distinguish between options and make conscious choices. It is necessary to realise that it pertains only to creating conditions for development, and not to coercion and imposing certain directions of such development. After all, to be autonomous or not is a choice of each person, and one cannot be forced to be autonomous [3, p. 408]. A similar position is shared by Esbjørn Melkevik and Bjarne Melkevik, who derive the human right to be master of oneself from the Kantian theory of dignity as autonomy. They consider dignity to be a matter of self-determination, and any excessive legal regulation, including that of morality, to be an obstacle to such self-determination. From this perspective, autonomy is considered the basis of human dignity, which once again confirms its fundamental status [16].

The second component of conditions of autonomy is the availability of an adequate range of options available to each person, between which one can make a choice [3, p. 369; 5, p. 25]. This condition directly depends on the level of civilisation development. From this it follows that autonomy is not a certain "natural state" but a civilisational achievement. Furthermore, autonomy is a "matter of degree", meaning that there may be more or fewer conditions for it in a society [3, p. 373], and the government's task is to ensure that these conditions are expanded. However, it is not a question of a mechanical increase in the number of options, but of genuine diversity. Therefore, we can evaluate the level of autonomy by using the "diversity test" [3, p. 375]. The third component of autonomy is independence, i.e. the absence of unjustified coercion and manipulation. In this respect, Ratz directly equates manipulation with coercion, which is particularly interesting in the context of our study [3, p. 378, 421].

This makes clear the role of personal autonomy within the legal system. Autonomy is not only directly related to and serves as the foundation of the fundamental values of dignity and freedom, and thus the foundation of human rights. It is also a prerequisite for a person to be considered a subject of law

and not an object. As Berlin notes, "All forms of tampering with human beings, getting at them, shaping them against their will to your own pattern, all thought-control and conditioning, is, therefore, a denial of that in men which makes them men and their values ultimate" [7, p. 184]. Without a certain level of autonomy, we cannot speak of a person's free will, and therefore of his or her ability to enter into transactions [12]. However, the foundations of modern legislation were laid at a time when any actual threats to personal autonomy were much more visible and much less sophisticated than now. Consequently, they could be overcome with the help of an established list of rights guaranteeing human freedom, without direct reference to autonomy or the conditions necessary for its existence.

Indeed, if we look at Stalin's USSR or Orwell's dystopia [17], we will see that there is neither freedom nor autonomy, but only coercion and violence. And so it may seem that autonomy and freedom are identical. Conversely, if we consider Huxley's "Brave New World" [18], that is clearly focused on future threats rather than contemporary ones, the difference between autonomy and freedom becomes clear [5, p. 25]. In Huxley's dystopia, there is virtually no direct violence and a relatively high level of freedom is provided (at least from the point of view of the characters). When the "Savage" John enters this world, he can do and say almost anything there. Meanwhile, autonomy in that place is minimised through sophisticated manipulation, and it is the impossibility of personal autonomy, even in a remote place, that drives this character to suicide.

The impact of artificial intelligence on personal autonomy

The processes described below, that negatively affect PA, started even before the widespread introduction of AI algorithms. AI can significantly intensify them and bring them to a new level, which amplifies the threats to PA. The situation is complicated by the fact that the decrease of autonomy caused by the introduction of AI is difficult to discern, record and prove, which is critical for the protection of individuals' rights. Autonomy is almost impossible to quantify and measure in order to remedy the problem with the help of algorithms [11, p. 3]. In addition, many technological solutions in the field of influencing people's opinions and behaviour are owned by companies that invest heavily in these developments and are not publicly available for analysis [12, p. 8]. Nevertheless, it can hardly be denied that the impact of AI algorithms on PA is significant.

The most obvious direction of such impact is the manipulation of a person, in particular the adaptive preference formation [11, p. 9]. As Eliza Mik notes, such manipulations towards reducing PA are not a single technology, but a

whole complex of many elements that mutually reinforce each other [12, p. 6]. They accompany users at all stages of using online services. On the present-day Internet, a whole "ecosystem" of "ambient and pervasive manipulations" has been formed [12, p. 6], which cannot be left if you continue to use common services. Ways of influencing people include manipulations of attention, choice, feelings, states, environment, design, and much more. Smart algorithms allow fine-tuning the parameters for each individual, even without using or storing one's personal data. Human behaviour and decision-making when using websites are analysed at the neurological level to achieve maximum effect [12, p. 15]. Companies like Neurons AI promise their clients to considerably improve their marketing results with the help of AI and based on the latest advances in neuroscience. Large platforms are designed to exploit all known cognitive biases of users. A fatal disproportion arises when the online platform or advertiser knows much more about the user than the user himself [12, p. 12-13].

It is necessary to realise that all these manipulations are carried out purposefully in the interests of specific actors – usually for commercial benefit, but also for political and other gains. The most well-known example of political manipulation relates to Cambridge Analytica that influenced political processes and elections based on psychological targeting [19]. Nowadays, few people are surprised by the amazing success of high-tech election campaigns, which can result in the unexpected entry of random figures "from TikTok" into the final of the presidential election in a large country [20].

The second direction of influence on humans, which leads to a reduction of their autonomy, is the distortion of their perception in matters related to AI. A striking example of this is the idea imposed on society that AI is akin to humans, with the prospect of turning it into a kind of superhuman. Thus, algorithms are said to "think", "recognise", "see", "understand", "draw conclusions" etc. The very name "artificial intelligence" generates associations with human intelligence, and in the worst case scenario, AI is considered to be "true" intelligence, and human intelligence is seen as a primitive and imperfect prototype. Erroneous or false outputs from AI due to imperfect algorithms or insufficient data for "training" (another misleading term) are called "AI hallucinations" [21] as if it has a psyche and can be sick.

Robotics professor Cindy Grimm argues that this approach is harmful as it creates false assumptions and expectations and does not contribute to an adequate understanding of human-made technologies that would help people make informed decisions. She proves that the use of the term "black box" in relation to these technologies is unjustified, giving them an aura of grandeur and

mystery, since a modern "AI or robotic system is still far less complex than the average bacterium" and it "will just 'do' what you want" [22], or rather what its owner company wants. With this in mind, Grimm calls for the use of mechanistic and pedantic language when talking about AI. However, in reality, companies benefit from spreading such myths about AI, as it fuels interest in the subject, increases profits and investments, and allows them to shift attention from human autonomy to the autonomy of AI systems, allegedly capable of becoming an independent and uncontrollable agent (which in turn partially relieves owners and operators of responsibility and justifies a strict regime of AI development only in the hands of large companies or under government control). All of the above may give rise to excessive enthusiasm and trust in AI, which is supposedly not as biased as humans and to which one can transfer one's autonomy [11, p. 5]. In fact, it is the interests of owners and operators that are behind all this. A recent study in cognitive science demonstrates how human interaction with AI mutually reinforces biases [23]. In contrast, no such reinforcement occurs in human-to-human interaction, which may indicate that people uncritically accept the AI model outputs. A similar but older problem is the uncritical acceptance of the results provided by Google in response to a query and the perception of them as higher wisdom ("Google knows everything") or a mirror image of reality with complete disregard for the fact that the company pursues commercial and other interests and its algorithms are not transparent [12, p. 16-17].

As the above study proves [23] very minor initial biases on the part of humans or AI can be significantly amplified by a positive feedback loop and produce a considerable effect in the end. Even more threatening is the deliberate use of such biases by the owner companies, which can completely disorient people, cause them to gradually lose their autonomy, and ultimately turn them into a resource that businesses and governments are fighting over. At the same time, the personalisation of websites, along with the fine-tuning of AI to the user's mental characteristics and states, creates a false sense of comfort and trust and does not contribute to an adequate assessment of the situation and to a rational choice from the range of available options.

The third direction of influence on autonomy concerns the online architecture. A well-known Stanford law professor Lawrence Lessig identifies four modalities that regulate behaviour: law, social norms, the market, and architecture [24]. On the Internet, regulation by architecture is carried out mainly through computer code, including AI algorithms. Companies owning major Internet platforms are able to promote their own version of architectural regulation by narrowing the range of options available to the user, but hiding it through the provision of many similar options or the availability of insignificant settings that

can supposedly increase or decrease the protection of user rights. These trends are particularly evident in the field of personal data, where the data subject's consent to the processing of data or the user's consent to accept cookies has become an extremely annoying and empty formality which has virtually no impact on increasing autonomy, protecting the human rights or strengthening the position of people, but instead serves as a cover for business activities and government inefficiency [5, p. 40]. Nowadays, more and more experts recognise the inadequacy of the consent mechanism, but instead they propose to replace it with cooperation between the government and big business aimed at assessing risks and mitigating them [25]. In other words, they propose to completely remove humans from the decision-making process, as modern digital reality is supposedly too complex for them to understand. This solution could be catastrophic for PA, as it completely ignores all three conditions of autonomy, in particular the development of human mental capacities, the availability of various options and alternatives, and the possibility of choosing between them.

The introduction of AI can exacerbate these negative trends, as its settings directly depend on the owning companies and their interests. Modern large language models require huge resources for their training and operation, which will contribute to further monopolisation and concentration of power in the hands of the major players. Big companies themselves are demanding that AI development be restricted by law to everyone else, which is intended to guarantee their exclusive position in the market [26]. AI can easily and endlessly multiply similar products and content and distribute them online, thereby pushing real alternatives far to the periphery, where they will be found only by those who know exactly what to look for, and filling all available space with fake diversity generated by several similar AI models based on the same data from the past. We can already see evidence of the disappearance of diversity caused by AI. For example, the use of certain phrases in research articles or reviews has increased tens of times in recent years due to the use of AI by authors [27]. The process of generating results by AI models is even less transparent than searching with Google. The user is completely unaware of the sources of information or data used for training, and it is very difficult to judge the validity or objectivity of the result. The transition of professionals, experts, and students from searching for information in specialised databases or even via Google to ready-made answers provided by large language models can significantly narrow their field of vision, weaken critical thinking, reduce the diversity of input data, and thus negatively affect their autonomy. Furthermore, it opens up limitless possibilities for subtle manipulations by AI owners and operators that are difficult to detect and control.

Along with establishing the architecture, large companies push social norms beneficial to them that legitimise their business practices and encourage users to participate. Their strong regulatory potential is based on the huge power imbalance between online platform companies and people, most of whom can hardly be called their customers, but rather resources, while their real customers, such as advertisers, share much the same interest in reducing the PA of individuals [28].

Thus, it can be concluded that the introduction of AI in its current form can have a significant negative impact on all three groups of PA conditions. Firstly, nothing is being done to help people in the general public gain a better understanding of AI and improve their mental capacity to navigate the situation. On the contrary, myths about AI as an incomprehensible entity with inexhaustible potential are being reinforced, which disorient people and discourage them from dealing with complex technical aspects. Secondly, AI in its current form does not contribute to the increase of genuine diversity, since trends in this area are set by the same monopoly companies that pursue their narrow interests and try to channel people's choices in a way that is favourable to them. And thirdly, AI significantly enhances the manipulative capabilities of companies as well as political authorities, and renders ordinary users powerless. Under such conditions, individuals are increasingly becoming an object for high-tech manipulations and a resource that businesses and political actors want to control.

The issue of personal autonomy in AI regulations

Respect for personal autonomy as a key principle is mentioned in some declarative documents on AI regulation [11, p. 2]. In particular, respect for human autonomy is the first of four ethical principles, rooted in fundamental rights, set out in the European Commission's High-Level Expert Group's Ethical Guidelines for Trustworthy AI [29]. Among other things, the document states that respect for human autonomy "is strongly associated with the right to human dignity and liberty" and that "AI systems should not unjustifiably subjugate, coerce, deceive, manipulate, condition or herd humans" and "should be designed to augment, complement and empower human cognitive and cultural skills" [29, p. 11-12]. It also mentions such problems as illegitimate coercion, threats to mental autonomy, deception and unfair manipulation [29, p. 14].

The final document of the Montreal forum on the socially responsible development of AI, entitled Montreal Declaration for Responsible AI, also includes the principle of respect for autonomy as the second of the ten key principles. Importantly, the document elaborates on this principle, and its

interpretation largely coincides with the previously mentioned theoretical concepts of PA. In particular, it recognizes the right of individuals "to achieve their goals and live in accordance with their values and ethical beliefs" and the obligation of governments and companies not to promote or discredit "a certain conception of the good life" with the help of AI. It also recommends "to empower citizens ... by ensuring access to relevant forms of knowledge, promoting the learning of fundamental skills and encouraging the development of critical thinking" [30]. In general, the analysis of the Declaration shows that it reflects the need to promote all three categories of conditions of PA discussed earlier.

While the above two documents embody a human-centric approach to AI regulation, some other important ones, such as the OECD AI Principles [31] and the European Commission's White Paper on AI, mention personal autonomy only once and pay much less attention to it.

The main mandatory legal instrument regulating AI issues, the significance of which extends beyond the EU, is the AI Act. It mentions PA several times in recitals, but not directly in the articles. Instead, the autonomy of AI is mentioned more times, including in articles, and it is emphasised that such systems "have some degree of independence of actions from human involvement and of capabilities to operate without human intervention" [1, rec. 12]. The autonomy of humans is addressed in two recitals. Recital 27 restates the already mentioned Ethics Guidelines for Trustworthy AI. Recital 29 refers to manipulative techniques that "can be used to persuade persons to engage in unwanted behaviours, or to deceive them by nudging them into decisions in a way that subverts and impairs their autonomy, decision-making and free choices". It also mentions the possibility of deploying "subliminal components such as audio, image, video stimuli that persons cannot perceive, as those stimuli are beyond human perception, or other manipulative or deceptive techniques" [1, rec. 29] and refers to Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market, which, among other things, is intended to protect against such practices. In the articles of the AI Act, the provision that reflects the substantial threats of AI is the prohibition of certain uses of AI contained in Article 5. Among other things, it is placing on the market, putting into service or using an AI system that "deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner

that causes or is reasonably likely to cause that person, another person or group of persons significant harm", as well as "exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation" [1, art. 5.1.a-b].

Obviously, this provision covers a small number of manipulations on the Internet aimed at reducing PA, and even fewer of them can be recorded and proved. It refers to the use of "subliminal techniques" and specific harm to individuals, but as noted earlier, the main harm to autonomy is manifested in something quite different from harming a specific individual, namely the destruction of the conditions of autonomy. In this regard, one of the assessments of the draft AI Act 2022 by the Future of Life Institute recommends removing the phrase "subliminal techniques" so that the ban applies to any manipulative techniques, since the "term 'subliminal' is not explicitly defined" and because most "uses of AI will not be subliminal since they will be consciously perceived by users" [32]. The document also recommends adding societal harm to the list of harms because certain AI systems "can cause only modest harm to individuals but hurt societies at large" [32]. However, the final version of the Regulation does not incorporate these recommendations.

To some extent, the Act's provisions on systemic risk reflect concerns about societal harm. Part V of the Regulation distinguishes between general-purpose AI models with systemic risk and all other general-purpose AI models [1]. The criteria for the designation of general-purpose AI models with systemic risk are specified in Annex XIII, but they are more related to technical parameters than to the characteristics of potential harm. As for the nature of systemic risks, recital 110 states that they include, but are not limited to, any actual or reasonably foreseeable negative effects in relation to public health and safety, democratic processes, public and economic security, the dissemination of illegal, false, or discriminatory content. Nothing in this list directly indicates threats to autonomy, but it can probably be supplemented by referring to Article 1 of the Regulation which states that its overall purpose is "protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems" [1]. Human rights can already be seen here, but there are doubts that their conventional interpretations cover all aspects of respect for personal autonomy.

To make sure of this, let us turn to the case law of the European Court of Human Rights (ECtHR), which directly concerns the EU and its member states as parties to the European Convention on Human Rights (ECHR). The ECtHR

recognises and protects the right to autonomy but primarily as an element of the right to respect for private and family life and in the context of certain interpretations of dignity [33]. It has to do so because the Court's goal is to protect the rights of an individual, and it has to rely on the applicable articles of the ECHR. However, in the current online environment, this may not be enough to protect PA and preserve its conditions. As for the AI Act, even if certain AI systems that pose threats to PA are recognised as having systemic risk, this will still not guarantee autonomy, as its fundamental importance is not mentioned anywhere in the Regulation, and the main responsibility for assessing such risks lies with the providers of AI systems, which creates a conflict of interests in case these providers are interested in reducing PA.

A fundamentally different approach is implemented in the Council of Europe Convention on AI. It enshrines respect for individual autonomy along with respect for dignity in a separate article as a key principle – the first in a list of seven principles that also include other human rights-related issues: equality and non-discrimination, privacy and protection of personal data [2, art. 7-13]. The preamble of the Convention states that "artificial intelligence systems may undermine human dignity and individual autonomy, human rights, democracy and the rule of law" and also mentions "arbitrary or unlawful surveillance and censorship practices that erode privacy and individual autonomy". In other words, the Convention is fully consistent with the mentioned theories of PA and enshrines respect for autonomy at a very fundamental level, alongside (not within) dignity and before human rights, democracy and the rule of law, which are possible only when a certain degree of PA is ensured. This is a sound basis for the development of specific rules on the protection of PA in the context of AI regulation, but much depends on further documents of the Council of Europe that will detail this principle as well as on the quality of its implementation at the national level.

Conclusions

Based on the study of theoretical developments, it can be concluded that personal autonomy plays a fundamental role in a legal system, as it is the basis for the subjectivity of individuals and a prerequisite for the realisation of human rights. When establishing PA, one should bear in mind that the primary importance is not to deal with specific violations or interferences with the autonomy of individuals but to create, maintain and develop conditions of autonomy, which include: certain mental capacities of individuals that allow them to be autonomous, the existence of an adequate list of options and alternatives, the absence of coercion and manipulation not justified in terms of ensuring

autonomy. PA does not exist in nature; it is a civilisation's achievement. The loss of autonomy will not necessarily destroy civilisation, but it can irreversibly change it and bring it closer to a Huxley's dystopia resulting in a complete erosion of human rights and related values.

AI can significantly intensify the negative trends in the shrinking of PA that have already emerged in the past decades under the influence of information technology. The current directions in the development and implementation of AI, driven by the commercial interests of leading companies and sometimes by the interests of authoritarian rulers, are often associated with the degradation of the conditions of PA. This includes the lack of public education about the nature and limitations of AI, the deliberate spread of unjustified myths about AI, the reduction of diversity through monopolisation, architectural regulation, information and power imbalances, the unjustified use of AI, and the significant increase in the manipulative capabilities of owners and operators of AI systems. This poses a real threat of the final objectification of people and further loss of their influence under conditions of high-tech manipulation by powerful economic and political actors.

In the present context, we need to constantly ask ourselves whether particular decisions and trends are conducive to the development of PA and the promotion of genuine diversity. If society is not aware of the importance of human autonomy, it will tend to make wrong decisions, sacrificing it for other benefits and advantages that the AI and other technologies supposedly promise. However, it is completely unjustified to consider AI itself a problem, as it has no will of its own, and the lines of its development are entirely set by people. The problem arises when the balance of rights and interests in society is destroyed, and the system tilts too much in one direction – towards satisfying the interests of large companies or technocratic power. This leads to the ignoring of the true deep interests of people and society.

The analysis of legal documents has revealed two approaches to AI regulation. The first one relegates PA to the periphery and suggests that problems should be solved through cooperation between government and business with the help of risk assessment by experts in particular fields. This should result in ready-made solutions that are offered to people. The second human-centred approach emphasises the protection of personal autonomy. However, the creation of detailed rules within this approach is a matter for the future. Their development requires considerable theoretical work, as the existing dominant concepts of human rights, democracy and the rule of law do not fully reflect the current threats in the online environment, which may be fundamentally different from

the real world and often do not take into account the urgent need to protect PA in the present-day situation.

References

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence. (June, 2024). Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

[2] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. (2024). Retrieved from: https://rm.coe.int/1680afae3c.

[3] Raz, J. (1986). *The morality of freedom*. Oxford: Clarendon.

[4] Dworkin, G. (1988). *The theory and practice of autonomy*. Cambridge University Press.

[5] Bernal, P. (2014). *Internet privacy rights: Rights to protect autonomy*. Cambridge University Press.

[6] Mill, J.S. (2003). *On Liberty*. Yale University Press.

[7] Berlin, I. (1969). *Four essays on liberty*. Oxford University Press.

[8] Risse, M. (2019). Human Rights and Artificial Intelligence: An Urgently Needed Agenda. *Human Rights Quarterly, 41*(1), 1-16.

[9] Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology, 4*, 100005. https://doi.org/10.1016/j.jrt.2020.100005.

[10] Shaelou, S.L., & Razmetaeva, Yu. (2024). Challenges to Fundamental Human Rights in the age of Artificial Intelligence Systems: shaping the digital legal order while upholding Rule of Law principles and European values. In *ERA Forum, 24*(4), 567-587. https://doi.org/10.1007/s12027-023-00777-2.

[11] Prunkl, C. (2024). Human autonomy at risk? An Analysis of the Challenges from AI. *Minds and Machines, 34*(3), 26. https://doi.org/10.1007/s11023-024-09665-1.

[12] Mik, E. (2016). The erosion of autonomy in online consumer transactions. *Law, Innovation and Technology*, *8*(1), 1-38.

[13] Solsman, J.E. (2018). YouTube's AI is the puppet master over most of what you watch. *CNET*. Retrieved from https://www.cnet.com/news/youtube-ces-2018-neal-mohan/.

[14] Sukhorolskyi, P. (2021). Right to be forgotten and freedom of expression: Peculiarities of balancing. In *Right to be Forgotten* (pp. 146-162). Kharkiv: ECUS.

[15] *Google v. Agencia Espacola de Protectiyn de Datos (AEPD) and Mario Costeja Gonzalez*, 13 May 2014, European Court of Justice, Case C-131/12.

[16] Melkevik, E., & Melkevik, B. (2021). Two concepts of dignity: On the decay of agency in law. In *The Inherence of Human Dignity. Vol. 1. Foundations of Human Dignity* (pp. 133-147). London: Anthem Press.

[17] Orwell, G. (1949). *Nineteen Eighty-Four*. London: Secker & Warburg.

[18] Huxley, A. (1932). *Brave New World*. London: Chatto & Windus.

[19] Carroll, D.R. (2021). Cambridge Analytica. In *Research Handbook on Political Propaganda* (pp. 41-50). Edward Elgar Publishing.

[20] Goury-Laffont, V. (December 21, 2024). Report ties Romanian liberals to TikTok campaign that fueled pro-Russia candidate. *Politico*. Retrieved from https://www.politico.eu/article/investigation-ties-romanian-liberals-tiktok-campaign-pro-russia-candidate-calin-georgescu/.

[21] Waldo, J., & Boussard, S. (2025). GPTs and Hallucination. *Communications of the ACM, 68*(1), 40-45. https://doi.org/10.1145/3703757.

[22] Grimm, C. (2021). The danger of anthropomorphic language in robotic AI systems. *The Brookings Institution*. Retrieved from https://www.brookings.edu/articles/the-danger-of-anthropomorphic-language-in-robotic-ai-systems/.

[23] Glickman, M., & Sharot, T. (2024). How human – AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour,* 1-15.

[24] Lessig, L. (1999). *Code and Other Laws of Cyberspace*. New York: Basic Books.

[25] Mantelero, A. (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review, 34*(4), 754-772. https://doi.org/10.1016/j.clsr.2018.05.017.

[26] Chakravorti, B. (2024, January). What if regulation makes the AI monopoly worse? *Foreign Policy*. Retrieved from https://foreignpolicy.com/2024/01/25/ai-regulation-monopoly-chatgpt/.

[27] Shapira, P. (2024). *Delving into "delve"*. Retrieved from https://pshapira.net/2024/03/31/delving-into-delve/.

[28] Sukhorolskyi, P. (2024). Fundamental values of data protection law: Autonomy vs the Megamachine. In *European Fundamental Values in the Digital Era*. Kharkiv: Pravo.

[29] High-level expert group on artificial intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission.

[30] Dilhac, M.A., Abrassart, C., & Voarino, N. (2018). *Report of the Montréal Declaration for a responsible development of artificial intelligence*. Retrieved from https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/27795/UdeM_Decl_IA-Resp_Rapport_complet_EN.pdf.

[31] OECD. (May, 2019). *Recommendation of the Council on Artificial Intelligence*. Retrieved from https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf.

[32] Uuk, R. (January, 2022). *Manipulation and the AI Act*. The Future of Life Institute.

[33] Koffeman, N.R. (2010). *(The right to) personal autonomy in the case law of the European Court of Human Rights*. Leiden: Leiden University. Retrieved from https://hdl.handle.net/1887/15890.

Petro M. Sukhorolskyi
Ph.D. in Law, Associate Professor
Associate Professor of the Department of Political Science and International Relations
Lviv Polytechnic National University
79000, 12, Stepan Bandera Str., Lviv, Ukraine
e-mail: petro.m.sukhorolskyi@lpnu.ua
ORCID 0000-0002-1689-3283

Петро Михайлович Сухорольський
кандидат юридичних наук, доцент
доцент кафедри політології та міжнародних відносин
Національний університет «Львівська політехніка»
79000, вул. Степана Бандери, 12, Львів, Україна
e-mail: petro.m.sukhorolskyi@lpnu.ua
ORCID 0000-0002-1689-3283