ЦИВІЛЬНЕ ПРАВО І ЦИВІЛЬНИЙ ПРОЦЕС; СІМЕЙНЕ ПРАВО; МІЖНАРОДНЕ ПРИВАТНЕ ПРАВО

UDC 347.77/.78
DOI https://doi.org/10.32782/chern.v1.2024.6

D. P. Bohatchuk
PhD in Law (Candidate of Science of Law),
Senior Lecturer of the Faculty of Law
National University of "Kyiv-Mohyla Academy"
orcid.org/0009-0008-6804-6765

INTELLECTUAL PROPERTY LAW TO PROTECT ARTIFICIAL INTELLIGENCE SYSTEMS AGAINST ADVERSARIAL ATTACKS IN LIGHT OF THE PURPOSES OF INTELLECTUAL PROPERTY LAW[1]

The paper analyzes whether the application of intellectual property law as a countermeasure against adversarial attacks on the artificial intelligence systems is consistent with the purposes of the intellectual property law.

In view of rapid development of artificial intelligence and growing number of domains in which artificial intelligence systems are applied, the adversarial attacks on such systems can potentially cause very harmful effects. Thus, there is a need to identify and develop countermeasures against adversarial attacks, in particular with the help of law. The issue of the possible legal measures that can be applied to counter adversarial attacks has to be considered first of all in the light of the purposes of the law. This paper is dedicated to an overview of the purposes and theories of justification of intellectual property law, in particular, copyright and patent law, which are further considered in the context of artificial intelligence and adversarial attacks. Although the theoretical foundations underlying the intellectual property system is a subject matter of debates, the purpose of incentivizing creativity and innovation could be considered as the primary purpose of intellectual property law.

The author concludes that the application of intellectual property law to protect AI systems against adversarial attacks and further integration of artificial intelligence into the intellectual property law system may be conducive to achieving the purposes of this system and implementing transparency. At the same time, modern intellectual property law should be adapted to the present-day realities, including the development of artificial intelligence.

This paper can serve as a basis for further research into the possibilities and expediency of the application of intellectual property law to protect artificial intelligence systems and to counter adversarial attacks on artificial intelligence.

Key words: adversarial attacks, artificial intelligence, intellectual property law, purposes of law.

Богатчук Д. П. **ПРАВО ІНТЕЛЕКТУАЛЬНОЇ ВЛАСНОСТІ ДЛЯ ЗАХИСТУ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ ВІД «ЗМАГАЛЬНИХ АТАК» (ADVERSARIAL ATTACKS) У СВІТЛІ ЦІЛЕЙ ПРАВА ІНТЕЛЕКТУАЛЬНОЇ ВЛАСНОСТІ**

У статті проаналізовано, чи відповідає застосування права інтелектуальної власності для протидії «змагальним атакам» (adversarial attacks) на системи штучного інтелекту цілям системи права інтелектуальної власності.

Зважаючи на стрімкий розвиток штучного інтелекту та зростаючу кількість напрямків, в яких застосовуються системи штучного інтелекту, «змагальні атаки» (adversarial attacks) на такі системи потенційно можуть спричинити дуже шкідливі наслідки. Таким чином, виникає потреба у визначенні та розвитку заходів протидії «змагальним атакам» (adversarial attacks), зокрема, за допомогою права. Питання про можливі механізми правового захисту, які можуть бути застосовані для протидії «змагальним атакам» (adversarial attacks), слід розглядати, перш за все, у світлі цілей права. Ця стаття присвячена огляду цілей і теорій щодо обґрунтування права інтелектуальної власності, зокрема, авторського права і патентного права, які в подальшому розглядаються в контексті штучного інтелекту і «змагальних атак» (adversarial attacks). Хоча теоретичне підґрунтя системи інтелектуальної власності є предметом дебатів, основною ціллю права інтелектуальної власності можна вважати стимулювання творчості та інновацій.

Автор робить висновок, що застосування права інтелектуальної власності для захисту систем штучного інтелекту від «змагальних атак» (adversarial attacks) та подальша інтеграція штучного інтелекту в систему права інтелектуальної власності може сприяти досягненню цілей цієї системи та впровадженню прозорості. Водночас, сучасне право інтелектуальної власності має бути адаптоване до реалій сьогодення, в тому числі до розвитку штучного інтелекту.

Ця стаття може слугувати основою для подальших досліджень щодо можливостей та доцільності застосування права інтелектуальної власності для протидії «змагальним атакам» (adversarial attacks) на системи штучного інтелекту.

Ключові слова: adversarial attacks, штучний інтелект, право інтелектуальної власності, цілі права.

[1] The author thanks Prof. Dr. Alfred Früh, University of Basel, Faculty of Law, the Center for Life Sciences Law (ZLSR), for his great support during the author's research at the Center for Life Sciences Law (ZLSR), Faculty of Law, the University of Basel. This paper represents the views of the author only.

Introduction. Taking into consideration the rapid development, growing impact, and widespread use of the artificial intelligence systems (AI systems), the study of adversarial attacks on AI systems requires substantial scientific attention. Adversarial attacks on AI have been the subject of research by scholars in the field of technology, including Raphaël Dang-Nhu, Mahmood Sharif, Ian J. Goodfellow, and others. At the same time, the legal study of adversarial attacks on AI systems is very limited, and the topic has been studied by very few legal scholars, in particular by Alfred Früh and Dario Haux. Despite many legal questions and huge legal problems caused by adversarial attacks on AI, many problems in this area remain without comprehensive answers and complex solutions. The foundations and justification of intellectual property (IP) law, patent law, copyright law have been studied by many scholars and experts in the field of IP: Fritz Machlup, Lionel Bently, Brad Sherman, Dominique Guellec, Henrik Timmann, Maximilian Haedicke, Jennifer Davis, and many others. Scholars Reto M. Hilty, Jörg Hoffman, and Stefan Scheuerer have devoted their research, in particular, to the issue of IP justification for artificial intelligence. This paper is intended to be an introductory work which shall help to find out whether application of IP law as one of the tools for protection of AI systems against adversarial attacks aligns with the purposes of IP law and, therefore, whether further research on possible means of IP law for the protection of AI systems could be helpful for building an effective system of legal protection against adversarial attacks.

Main body. AI system can be defined as "a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" [1]. Adversarial attacks on AI systems use the perturbations to the input data [2, p. 1] to misdirect the AI systems and to lead to incorrect outputs such as misclassifications or false predictions [3, p. 1, 4].

An example of adversarial attacks is the implementation of so-called "smart silencers" to prevent the identification of the class and caliber of a gun which fired a shot [4, p. 41].

By adding perturbations to the input data, an attacker can cause a target AI model to produce completely wrong outputs [5, p. 2]. Adversarial attacks are not limited to AI systems that work with classification and can also target other AI systems and cause them, for example, to make wrong predictions [6, p. 4, 6, 8].

Adversarial attacks can potentially cause harmful effects in various domains in which AI systems are applied. Therefore, there is an urgent need to identify countermeasures that can prevent or remedy such attacks.

Computer science has proposed several measures to counter adversarial attacks on a technical level [3, p. 13-16] and new research is constantly being published in this field.

However, the issue of countering adversarial attacks with the help of law requires a complex study, first of all, in view of the purposes of law as a system of regulation of social relations.

The issues which arise within the IP system should be solved, depending on the purposes this system is supposed to serve [7, p. 7]. We therefore need to consider whether mobilizing IP protection against adversarial attacks can be aligned with the purposes of IP law [3, p. 18]. This requires a closer look at such purposes and theories of justification, particularly for copyright and patent law, which will be covered by the scope of this paper.

There are different theories of justification of copyright and the restrictions imposed by it. These theories base, in particular, on the following arguments: (1) natural rights arguments, according to which copyright is a property right granted because the respective intellectual productions emanate from the mind of the authors, reflect individuality of their creators or constitute the products of the intellectual labour [8, p. 35, 36]; in this view, the author possesses a natural right to ownership of the results of the labour [9, p. 23]; (2) reward arguments, perceiving copyright as a fair reward and gratitude to an author for the effort in creating a work and sharing it with the public [8, p. 36, 37]; this reward may be proportional to the public's appreciation of the work, as the more copies of the work are purchased / listened, etc., the greater is the financial benefits of the copyright owner [8, p. 36, 37]; (3) incentive theories, arguing that copyright is an incentive for the creation and dissemination of works [8, p. 37]; it is said that, without copyright protection and profit therefrom, authors will be deprived of economic incentive to create and disseminate their works [9, p. 23]; (4) neo-classical economics arguments, which consider copyright as a promotor of optimal use of intellectual resources and means of prevention of over-exploitation [8, p. 38]. There are also many other justification theories and points of view on copyright.

All the theories of justification for copyright may be challenged [9, p. 23]. At that, the above-mentioned arguments can complement each other. The InfoSoc Directive proposes rather broad perception of the purpose of copyright, combining different approaches: "A rigorous, effective system for the protection of copyright and related rights is one of the main ways of ensuring that European cultural creativity and production receive the necessary resources and of safeguarding the independence and dignity of artistic creators and performers" [10]. Stimulating creativity can be seen as a primary purpose of copyright from the perspective of social value and public interest. Other goals of copyright (such as rewarding or recognizing) may be founded on this primary purpose.

The theories of justification of patent law are similar to those regarding copyright. There are two basic approaches to justification of patent system – the natural rights and utilitarian theories [11, p. 46].

The "natural-law" thesis assumes that person has the natural property rights in his own ideas, which rights should be recognized and protected, in particular, from the unauthorized use [7, p. 21]. According to the natural rights approach, the inventor, like any other worker, has a natural property right to the results of his or her labour [11, p. 46]. The "reward-by-monopoly" thesis, which also belongs to the natural rights approach, assumes that a person should receive reward for his services in proportion to their usefulness for the society [7, p. 21]. According to this approach, a patent can be considered a reward, which has been deserved by the inventor, rather than an incentive that serves to public interest [11, p. 47].

The economic approach to justification of the patent system is utilitarian: it sees patents as a policy instrument that functions in the interests of the society [12, p. 3] and that is connected with certain aims and circumstances [11, p. 49]. Pursuant to the "exchange-for-secrets" thesis [7, p. 21], which is one of the variations of the utilitarian argument, patent constitutes a contract between the inventor and society, by which society grants transitory monopoly to the inventor in exchange for patent disclosure [11, p. 50]. Thus, patents incentivize the disclosure of the knowledge by inventors to the benefit of society [11, p. 74] and "for the use of future generations" [7, p. 21]. However, disclosure comes after the invention is made, so having the invention made is the primary public interest and disclosure comes second [11, p. 51]. In view of this, patents can be seen as a promise of society to grant investors some exclusive right if they come with an invention, as a very special type of "contract" providing incentives to invent [11, p. 51]. According to the "monopoly-profit-incentive" thesis, inventions will not be obtained in sufficient measure if inventors and investors do not have profits connected with the competitive exploitation of technical knowledge [7, p. 21]. The thesis that the patent system produces effective incentives for inventing and stimulates technological progress is often considered as the fundamental economic justification for patents [7, p. 33]. Fostering innovation and growth [12, p. 3], encouraging the diffusion of technology by the economic mechanisms may be regarded as a general purpose of the patent system [12, p. 3; 13, p. 42]. Patent law aims for spurring innovation so as to promote technical progress by rewarding inventors with an exclusive market position through patent rights for a limited number of years [14, p. 5]. At the same time, patents incentivize sharing technical knowledge with the public instead of keeping it secret [14, p. 5]. In turn, patent disclosure facilitates follow up inventions and the invention of substitutes to the initial invention [11, p. 75].

The above-mentioned theories of justification of IP rights are only a few of many, and the debates over the paradigms underlying the IP system are indeed broad and controversial. In general, it is agreed that the purpose of the IP system as a whole is promoting innovation and creativity [15, p. 214]. Thus, according to the economic argument, the IP rights anticipate incentives for creation of new intellectual capital, without which incentives the respective subjects will not invest resources in creation of such intellectual capital [9, p. 5]. According to the rights-based approach, that rest on the ideas of the eighteenth-century philosopher John Locke, and according to the so called "labour" justification, the IP rights incentivize placing the creations before the public and, therefore, engender new ideas and further creativity [9, p. 6, 7]. At the same time, the assumption that the IP rights a priori provide incentives for innovation and creativity should be considered as false [16, p. 50]. In order to properly fulfill the functions and purposes of the IP rights, IP law must be adapted to the present reality, especially to the rise of AI technologies, that in turn affects the theories of justification of IP system, which are anthropocentric in their essence.

Thus, on the one hand, AI may cause a decrease in human effort to generate intangible goods, on the other hand, AI impacts market conditions and the level of human involvement in the creation and innovation process [16, p. 50-51, 53, 55].

Taken into account that strengthening IP protection of AI systems against adversarial attacks will be favorable for labor and financial investments in these systems and, thus, will encourage their further progress, mobilizing IP law as a countermeasure against adversarial attacks is generally consistent with the purpose of incentivizing creativity and innovation as the primary purpose of the IP law. However, this issue can also be viewed from a different angle: if the IP protection indeed has incentivizing effect on developments of AI, such fostered AI may leave people in disadvantageous competitive situation and, thus, pose a threat to human progress as a general purpose [16, p. 57; 17, p. 2053, 2106]. There is also another aspect to pay attention in this context: the self-regulatory sharing schemes of "Open Source" movement, etc., are in fact playing a crucial role in incentivizing AI innovations now and, at the same time, call into question IP law justification in realm of AI [16, p. 63, 64]. Moreover, modern technical measures, which keep constantly improving, can provide for protection for AI systems against adversarial attacks and currently play a major role for their security.

Thus, the above issue of interrelation between the IP law protection and AI systems is complicated indeed and requires a reference to the foundations to find out the proper legal response. The best options for the legal regulation in respect of AI should be determined by the legislators and policymakers in accordance with fundamental legal principles, in particular the rule of law [18]. As proceeds from the Report on intellectual property rights for the development of artificial intelligence technologies (Report A9-0176/2020), the European Parliament sees a need to strengthen IP protection for AI [19]. In this report the European Parliament “stresses the need for the Commission to aim to provide balanced and innovation-driven protection of intellectual property, for the benefit of European AI developers, to strengthen the international competitiveness of European companies, including against possible abusive litigation tactics, and to ensure maximum legal certainty for users, notably in international negotiations, in particular as regards the ongoing discussions on AI and data revolution under the auspices of WIPO” [19].

The well-established and international nature of the IP system serves as an argument to use this system for promotion of knowledge sharing of AI technologies [15, p. 222]. In particular, the benefits of patent protection, together with the legal requirements on sufficient patent disclosure could be an additional incentive for innovators to invest in so-called “explainable” and transparent AI [20, p. 217]. The need for transparency in AI, creating AI models that are able to explain themselves, or to make decisions that can be explained to people is based on the social and ethical reasons [21, p. 297]. Thus, IP protection of AI systems can be beneficial for the implementation of transparency as a legal value in this field. At the same time, wider knowledge and information sharing regarding AI within the IP system should also include the mechanisms of IP protection of AI against the potential infringements such as the adversarial attacks, as well as the IP means of decreasing the vulnerability of AI. In such a combination, the integration of AI in the IP system could be effective and socially beneficial.

Conclusion. The application of IP law to protect AI systems against adversarial attacks, and the further integration of AI into the IP system, may be conducive to achieving the purposes of IP law and implementing transparency under the fundamental principles of law.

Bibliography

1. Interinstitutional File 2021/0106(COD), Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 5662/24, Brussels (26 January 2024), Article 3. URL: https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf.

2. Ian J. Goodfellow, Jonathon Shlens and Christian Szegedy, ‘Explaining and Harnessing Adversarial Examples’ (20 March 2015) published as a conference paper at ICLR 2015. URL: https://arxiv.org/pdf/1412.6572.pdf.

3. Alfred Früh and Dario Haux, ‘Countermeasures against Adversarial Attacks on Computational Law’ (2024) Journal of Cross-disciplinary Research in Computational Law, CRCL, volume 2, issue 1. URL: https://journalcrcl.org/crcl/article/view/31/20.

4. Marcus Comiter, ‘Attacking Artificial Intelligence: AI’s Security Vulnerability and What Policymakers Can Do About It’ (2019) Harvard Kennedy School, Belfer Center for Science and International Affairs. URL: https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf.

5. Yao Deng, Xi Zheng, Tianyi Zhang, Chen Chen, Guannan Lou, Miryung Kim, ‘An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models’ (6 February 2020). URL: https://arxiv.org/pdf/2002.02175.pdf.

6. Raphaël Dang-Nhu, Gagandeep Singh, Pavol Bielik, Martin Vechev, ‘Adversarial Attacks on Probabilistic Autoregressive Forecasting Models’ (2020) Proceedings of the 37 th International Conference on Machine Learning, Vienna, Austria. URL: https://files.sri.inf.ethz.ch/website/papers/raphaelicml.pdf.

7. Fritz Machlup, *An Economic Review of the Patent System*, Study of the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, United States Senate, Eighty-fifth Congress, second session, Study No. 15 (Washington: US Government Printing Office 1958). URL: https://cdn.mises.org/An%20Economic%20Review%20of%20the%20Patent%20System_Vol_3_3.pdf.

8. Lionel Bently and Brad Sherman, *Intellectual Property Law* (3rd edn., Oxford: Oxford University Press 2008).

9. Jennifer Davis, *Intellectual Property Law* (3rd edn., Oxford: Oxford University Press 2008).

10. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167 (2001). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex-%3A32001L0029.

11. Dominique Guellec, ‘Patents as an Incentive to Innovate’ in: Dominique Guellec and Bruno van Pottelsberghe de la Potterie (eds.), *The Economics of the European Patent System: IP Policy for Innovation and Competition* (Oxford University Press 2011).

12. Dominique Guellec, Bruno van Pottelsberghe de la Potterie, ‘Introduction’ in: Dominique Guellec and Bruno van Pottelsberghe de la Potterie (eds.), *The Economics of the European Patent System: IP Policy for Innovation and Competition* (Oxford University Press 2011).

13. Dominique Guellec, ‘Historical Insights’ in: Dominique Guellec and Bruno van Pottelsberghe de la Potterie (eds.), *The Economics of the European Patent System: IP Policy for Innovation and Competition* (Oxford University Press 2011).

14. Henrik Timmann and Maximilian Haedicke, *Patent Law: a Handbook on European and German Patent Law* (München: C H Beck 2014).

15. Rachel Free, ‘Intellectual Property’, in: Charles Kerrigan (ed.), *Artificial Intelligence : Law and Regulation* (Cheltenham, UK Edward Elgar Publishing 2022).

16. Reto M. Hilty, Jörg Hoffman, and Stefan Scheuerer, ‘Intellectual Property Justification for Artificial Intelligence’, in: Jyh-An Lee, Reto M. Hilty, and Kung-Chung Liu (eds.), *Artificial Intelligence and Intellectual Property* (1st edn., Oxford University Press 2021).

17. Daniel J. Gervais, ‘The Machine As Author’ (2019) Iowa Law Review, Volume 105, Vanderbilt Law Research Paper No. 19-35.

18. The principle of the Rule of Law, Doc. 11343, Report, Committee on Legal Affairs and Human Rights, Rapporteur: Mr. Erik Jurgens, Netherlands, Socialist Group (6 July 2007). URL: https://assembly.coe.int/nw/xml/XRef/X2H-Xref-ViewHTML.asp?FileID=11593.

19. European Parliament, Committee on Legal Affairs, ‘Report on intellectual property rights for the development of artificial intelligence technologies’, A9-0176/2020, Rapporteur: Stéphane Séjourné (2020). URL: https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.html.

20. Jozefi en Vanherpe ‘AI and IP: A Tale of Two Acronyms’, in: Jan de Bruyne and Cedric Vanleenhove (eds.), *Artificial Intelligence and the Law* (Cambridge: Intersentia 2021).

21. Matt Hervey, Virginia Driver and Tom Woodhouse, ‘Intellectual Property’, in: Matt Hervey and Matthew Lavy (eds.), *The Law of Artificial Intelligence* (1st edn., Sweet & Maxwell 2021).