DOMINIK BIERECKI, CHRISTOPHE GAIE, MIROSŁAW KARPIUK,
JEAN LANGLOIS-BERTHELOT

Creating Resilient Artificial Intelligence Systems. A Responsible Approach to Cybersecurity Risks

Abstract

In a computerized state, where the information society makes extensive use of digital services, artificial intelligence is important. To facilitate the use of cyberspace, AI systems need to be resilient to cyber threats that can not only limit the operation of such systems, but also take control of them. Ensuring cybersecurity requires the introduction and use of appropriate protective measures at the stage of creating such systems, implementing them, and using them. It should be emphasized that the use of new technologies, including artificial intelligence, must be done responsibly, so as not to expose oneself and other users to risks, including those that cause significant damage. The authors use the dogmatic-legal method to conduct an analysis of legal acts that normalize both artificial intelligence and cybersecurity issues. The theoretical-legal method was used to analyze the literature on both normative issues related to artificial intelligence and its potential, as well as obligations related to cybersecurity, including countering threats occurring in cyberspace.

KEYWORDS: cybersecurity, artificial intelligence, cyberspace, cyber threats

DOMINIK BIERECKI – associate professor, Pomeranian University in Słupsk (Poland), ORCID – 0000-0001-6993-3974, e-mail: dominik.bierecki@upsl.edu.pl
CHRISTOPHE GAIE – associate researcher, Laboratoire Magellan, Université Jean Moulin Lyon 3 (France), French Prime Minister Services, ORCID – 0000-0002-8252-5278, e-mail: christophe.gaie@gmail.com
MIROSŁAW KARPIUK – full professor, University of Warmia and Mazury in Olsztyn (Poland), ORCID – 0000-0001-7012-8999, e-mail: miroslaw.karpiuk@uwm.edu.pl
JEAN LANGLOIS-BERTHELOT – head of the Division of Operational Sciences and Development (France), chair of teachings, Center for Advanced Military Training, ORCID – 0009-0001-2397-5930, e-mail: jeanlanglois4@gmail.com

HTTPS://DOI.ORG/10.36128/PRIW.VI58.1499

1 Introduction

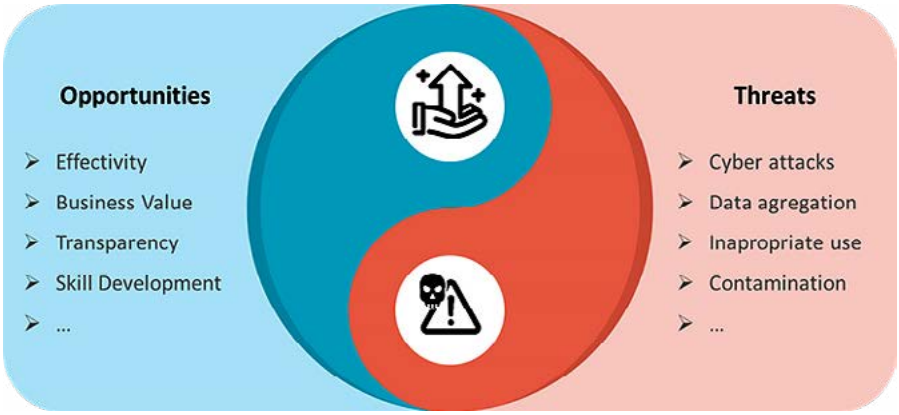
Artificial intelligence is everywhere today, used in public, business, and private life. Artificial intelligence systems facilitate not only information retrieval, business, education, work, and public tasks, but also everyday life. Indeed, according to a recent survey, the use of generative AI has risen from 65% to 71% in 2024.[1] Moreover, the usage of AI by organizations reaches very high levels in India (59%), the United Arab Emirates (58%), or China (50%), while European countries reveal a more prudent adoption, such as Spain (28%) and France (26%).[2] However, the adoption of AI is not only a convenience, but also a threat. Activities performed with these modern tools carry the risk of cyber threats, as demonstrated by the 614 AI incidents and hazards reported by reputable international media between November 2023 and January 2024.[3] On the one hand, those threats can lead to limiting or paralyzing the activities of artificial intelligence systems; on the other hand, they can be used for other harmful purposes. In the case of artificial intelligence, protection should involve preventing unauthorized access to its systems. Such access could allow an unauthorized entity to steal data or infect systems. Privacy and human rights and freedoms may be threatened as well. The benefits to be derived from the use of new technologies, including artificial intelligence, should not obscure the need to bear the costs associated with ensuring cybersecurity. Artificial intelligence operates in cyberspace and is therefore vulnerable to the threats that exist in cyberspace. The duality of AI that lies in its opportunities and threats is illustrated in Figure 1 below.

1 Alex Singla, Alexander Sukharevsky, Lareina Yee, Michael Chui, Bryce Hall, *The state of AI: How organizations are rewiring to capture value* (New York: McKinsey Global Publishing, 2025).

2 Katherine Haan, "22 Top AI Statistics and Trends. Forbes Advisor", https://www.forbes.com/advisor/business/ai-statistics/, [accessed: 21.06.2025].

3 "AI risks and incidents", https://www.oecd.org/en/topics/ai-risks-and-incidents.html, [accessed: 21.06.2025].

Figure 1: Duality of Artificial Intelligence



Education at various stages must play an important role in cybersecurity – this includes academic programs as well as ongoing professional development, which is important as we face the challenges of developing secure digital services. IT professionals should acquire the necessary skills to build robust and resilient ICT (Information and Communications Technology) systems. A continuous learning approach ensures that professionals are up to date with the latest threats and best practices in cybersecurity.[4] Digital education and public awareness are essential in increasing resilience to cyber attacks, which is critical in the context of the growing threats posed by technological advances. In the age of digitization, where technology permeates every aspect of life, the ability to use digital tools safely and with awareness is very important. Education in this area helps to understand the risks of cyber threats and teaches effective methods of countering them.[5]

Artificial intelligence is best used in a digital state and society, where both the public institutions themselves and their beneficiaries are digitally competent. Such competencies allow both effective and secure use of new technologies, including artificial intelligence.[6] In a digital state

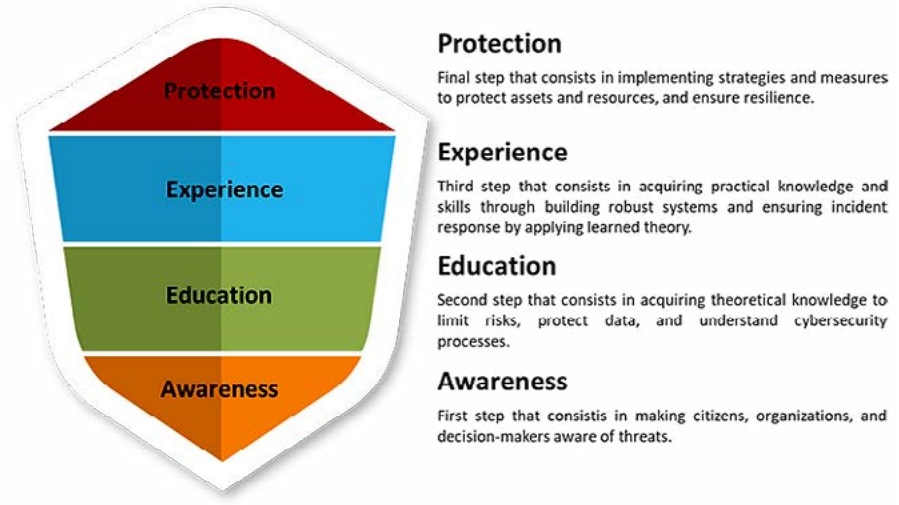
4 Christophe Gaie, Mirosław Karpiuk, Nicola Strizzolo, "Cybersecurity of Public Sector Institutions" *Prawo i Więź*, No. 6 (2024): 358.

5 Tomasz Wojciechowski, "Cyberbezpieczeństwo i dezinformacja we współczesnym świecie: strategie ochrony i zarządzania kryzysowego" *Ius et Securitas*, No. 1 (2024): 92.

6 Dominik Bierecki, Christophe Gaie, Mirosław Karpiuk, "Artificial Intelligence in e-Administration" *Prawo i Więź*, No. 1 (2025): 385.

and society with universal access to electronic services, cybersecurity is of particular importance, as it enables not only uninterrupted social communication, but also the proper functioning of strategic sectors of the economy.[7] Cybersecurity protects against threats occurring in cyberspace, and therefore ensures the normal operations of the state as a public entity on many levels.[8] Figure 2 below illustrates the different steps to set up an efficient cyber protection through the education of citizens, employees, IT specialists, and decision makers.

Figure 2: Different steps to establish a cyber protection through education



2 Artificial Intelligence and Cybersecurity

a. Definition of Cybersecurity

Cybersecurity is part of the general field of security, which is defined as a state in which not only individuals, but also communities, organizations, and states are adequately protected from threats to their well-being,

7 Jean Langlois-Berthelot, *Evaluating and Insuring Cyber Risks within Organizations* (Paris: EHESS, 2021).

8 Mirosław Karpiuk, "The Legal Status of Digital Service Providers in the Sphere of Cybersecurity" *Studia Iuridica Lublinensia*, No. 2 (2023): 190.

integrity, or survival. It includes protection against physical threats, as well as ensuring the economic, social, political and environmental conditions that allow for stable functioning.[9] Cybersecurity, as one of the fields of security, is concerned with preventing and responding to threats as well as removing their effects. It is essential not to forget to analyze the causes and sources of threats, as well as to apply security solutions against their occurrence. In the case of cybersecurity, the threats take place in cyberspace.[10] However, cybersecurity should be looked at from a multi-faceted perspective, which should take into account not only the ICT infrastructure or digital literacy, but also other factors such as the security environment or the international situation.[11] Fundamentally, cybersecurity aims to ensure the confidentiality, integrity, availability, and traceability of information and systems such as described in the EBIOS Risk Manager method.[12]

Cyber security, according to Article 2(4) of the Law of July 5, 2018 on the National Cyber Security System (Journal of Laws 2024, item 1077, as amended), is the resilience of information systems to actions that violate the confidentiality, integrity, availability, and authenticity of processed data or related services offered by these systems.[13] The national

---

9 Krzysztof Kaczmarek, "Wpływ zmian klimatycznych na bezpieczeństwo" *Journal of Modern Science*, No. 4 (2024): 412. For security issues, see also: Edyta Tkaczyk, "Bezpieczeństwo państwa w Konstytucji Rzeczypospolitej Polskiej. Refleksje nad dobrem chronionym" *Ius et Securitas*, No. 1 (2024): 42-50; Mirosław Karpiuk, "Glosa do wyroku Naczelnego Sądu Administracyjnego z dnia 12 lutego 2018 r. (II OSK 2524/17)" *Studia Iuridica Lublinensia*, No. 1 (2019): 185-194, doi: 10.17951/sil.2019.28.1.185-194; Jarosław Kostrubiec, Mirosław Karpiuk, Dominik Tyrawa, "The status of municipal government in the sphere of ecological security" *Hungarian Journal of Legal Studies*, No. 2 (2024): 164-181. https://doi.org/10.1556/2052.2024.00510.

10 Christophe Gaie, Mirosław Karpiuk, Andrea Spaziani, "Cybersecurity in France, Poland and Italy" *Studia Iuridica Lublinensia*, No. 1 (2025): 74.

11 Krzysztof Kaczmarek, Mirosław Karpiuk, Claudio Melchior, "A Holistic Approach to Cybersecurity and Data Protection in the Age of Artificial Intelligence and Big Data" *Prawo i Więź*, No. 3 (2024): 105-106.

12 "EBIOS Risk Manager – The method", https://cyber.gouv.fr/sites/default/files/2019/11/anssi-guide-ebios_risk_manager-en-v1.0.pdf, [accessed: 22.06.2025].

13 For a definition of cyber security, see also: Krzysztof Kaczmarek, "Finland in the light of cyber threats in the context of Russia's aggression against Ukraine" *Cybersecurity and Law*, No. 1 (2023): 212; Mirosław Karpiuk, "Recognizing an Entity as an Operator of Essential Services and Providing Cybersecurity at the National Level" *Prawo i Więź*, No. 4 (2022): 166-167; Krzysztof Kaczmarek, "Nordic countries in the face of digital threats" *Cybersecurity and Law*, No. 1 (2024): 152; Małgorzata Czuryk, "Cybersecurity and Protection of Critical Infrastructure" *Studia Iuridica Lublinensia*, No. 5 (2023): 44-45; Krzysztof Kaczmarek, "Vulnerability to cyber

cybersecurity system, according to Article 3 of this law, aims to ensure cybersecurity at the national level, including the uninterrupted provision of key services and digital services, by achieving an adequate level of security for information systems used to provide these services, and by ensuring the handling of incidents.

The European Union legislator defines cybersecurity as the activities that are necessary to protect networks and information systems, their users and others from cyber threats.[14] The domain is constantly evolving to address increasingly sophisticated threats and protect a growing digital landscape.

A synthesis of cybersecurity and cognitive security obligations has become essential to effectively address contemporary hybrid threats. In their manual on international law and cyberspace, Tsagourias and Buchan warned that hybrid threats exploit loopholes and normative gaps through attacks that produce strategic effects while remaining below the threshold of armed conflict. Although the European regulatory framework – notably the Joint Framework on Countering Hybrid Threats (JOIN(2016)18) – acknowledges this convergence, it still does not specify the legal obligations of private operators and AI system developers in this area.[15]

b. Definition of AI

According to the legal definition established by the European Union legislature, artificial intelligence should be understood as a machine system designed to operate with varying levels of autonomy, which can show adaptability once it is deployed and infer how to generate results affecting

threats: a qualitative analysis from societal and institutional perspectives" *Cybersecurity and Law*, No. 2 (2024): 108-109; Ewa Maria Włodyka, Krzysztof Kaczmarek, "Cyber Security of Electrical Grids – A Contribution to Research" *Cybersecurity and Law*, No. 2 (2024): 262-263.

14 Article 2(1) of Regulation (EU) 2019/881 of the European Parliament and of the Council of April 17, 2019 on ENISA (European Union Agency of Cybersecurity) and the certification of information and communication technology cybersecurity, and repealing Regulation (EU) No. 526/2013 (Cyber-Security Act) (OJ EU L 151, pp. 15-69).

15 Elias Tsagourias, James Buchan, *Research Handbook on International Law and Cyberspace* (Cheltenham: Edward Elgar Publishing, 2015); "Joint Framework on Countering Hybrid Threats (JOIN(2016)18)" (Brussels, 2016).

the physical or virtual environment based on input.[16] Artificial intelligence is an umbrella term that includes any type of software or hardware component that supports machine learning, computer vision, understanding, generation, and natural language processing, including robotics. Artificial intelligence enables machines to gather and analyze information about the environment and act toward a given goal. Based on observation and experience alone, some artificial intelligence systems will be able to adapt their behavior to the environment while acting autonomously.[17]

Establishing a clear and functional definition of AI requires considering the legal framework of the European Union and understanding its applicability in terms of development, deployment, and use. This definition determines which systems fall under specific legal obligations, such as those related to cybersecurity, risk management, and fundamental rights protection. This approach is necessary to ensure responsible AI adoption in areas like e-government and critical infrastructure.[18]

It seems necessary to explicitly integrate the concept of cognitive integrity into legal regimes related to AI and cybersecurity in order to fill this normative gap. This approach could draw inspiration from the guidelines of the Article 29 Working Party – now the European Data Protection Board. The Article 29 Working Party developed guidelines for automated decision-making (2018) and emphasized the importance of transparency and user understanding, focusing on the issue of preserving "informational self-determination."[19]

---

16 Article 3(1) of Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024 on establishing harmonized rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Official Journal of the EU L 2024/1689).

17 Krzysztof Kaczmarek, "Sztuczna inteligencja" [in:] *Leksykon cyberbezpieczeństwa*, ed. Katarzyna Chałubińska-Jentkiewicz (Warsaw: ASzWoj, 2024): 251-252. On artificial intelligence, see also: Tomasz Gergelewicz, "Bipolarity of Artificial Intelligence – Chances and Threats" *Ius et Securitas*, No. 2 (2024): 71-94.

18 Markus Mueck, Christophe Gaie, Dimitris Gkikas, "Introduction to the European Artificial Intelligence Act", [in:] *European Digital Regulations*, eds. Markus Mueck, Christophe Gaie (Cham: Springer, 2025): 53-90.

19 Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251rev.01).

In this sense, developers and operators of AI systems could be required to conduct a cognitive impact assessment, similar to the data protection impact assessments provided for in Article 35 of the GDPR. These assessments would aim to estimate the risks posed by AI in terms of influence, perceptions, or changes in user behavior. In a broad reading of Articles 13 and 14 of the GDPR, the implementation of accessible explainability mechanisms could be required, guaranteeing clear information on the logic and consequences of automated processing.

Finally, civil or administrative liability could be adapted to take into account cognitive harm, and in particular, as suggested by Calo (2018), in a logic of legal recognition of cyber manipulation as a new type of harm.[20][21]

c. Cybersecurity and the AI Act

The recitals of the Artificial Intelligence Act note the importance of ensuring the cybersecurity of high-risk artificial intelligence systems. Such systems should be covered, among other things, by cybersecurity requirements that are necessary to effectively mitigate risks to health, safety, and fundamental rights (recital 66). Recital 76 of the Artificial Intelligence Act states that cybersecurity plays a key role in ensuring that AI systems are immune to attempts by third parties acting in bad faith to modify their use, behavior, or performance, and to circumvent their safeguards by exploiting system vulnerabilities. The legal standard containing the order to ensure the cybersecurity of AI systems is expressed in Article 15(1) of the AI Act. According to this provision, high-risk AI systems should be designed and developed to achieve, among other things, an adequate level of cybersecurity, and to operate consistently in terms of cybersecurity throughout their life cycle. To this end, implementing in-depth security requires multiple countermeasures, for example, robust access controls, secure update mechanisms, and continuous monitoring for anomalies, which can help protect systems from various attacks. These countermeasures may prevent

[20] Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, May 4, 2016.

[21] Ryan Calo, "Artificial Intelligence Policy: A Primer and Roadmap" *U.C. Davis Law Review*, No. 2 (2018): 399-432.

attacks such as data poisoning, which aims to corrupt training data, or help avoid adversarial attacks where malicious inputs are designed to fool deployed models into generating erroneous decisions.[22]

In the recitals of the Artificial Intelligence Act, the European legislator also identifies the forms of cyber threats to AI. According to recital 76 of the Act, cyber attacks on AI systems may consist in exploiting specific AI assets, such as training datasets (e.g., data poisoning) or trained models (e.g., adversarial attacks or membership inference attacks), or in exploiting vulnerabilities in the digital assets of the AI system or the underlying ICT infrastructure. Further, the European legislator points out that providers of high-risk AI systems should implement appropriate measures, such as security controls, to ensure a level of cybersecurity commensurate with the risks described, taking into account, where appropriate, the underlying ICT infrastructure. Artificial intelligence functions based on data processed by ICT, which makes the cybersecurity of those systems crucial to AI cybersecurity.

Artificial intelligence systems expose users and organizations not only to technical but also to cognitive threats. The European Union Agency for Cybersecurity (ENISA, 2020) highlights how many attacks go beyond compromising data or models. Those attacks also aim to manipulate users' trust in the results generated by AI, and more broadly, in the information environment shaped by these systems.[23] The report by Brundage et al. (2018) further shows how the growing use of generative artificial intelligence in campaigns turns these technologies into hybrid vectors that combine cyber threats and information warfare.[24]

In its 2023 Global Risks Report, the World Economic Forum ranks AI-powered disinformation as a major risk for the coming decade. According to the report, this risk is more serious than purely technical cyber threats. However, current regulatory frameworks, such as the Cybersecurity Act (EU 2019/881) or the draft Regulation on Artificial Intelligence, are insufficiently equipped to meet the challenges of cognitive resilience. Cognitive

22 Shakila Zaman, Khaled Alhazmi, Mohammed Aseeri, Muhammad Raisuddin Ahmed, Risala Tasin Khan, Shamim Kaiser, Mufti Mahmud, "Security Threats and Artificial Intelligence based Countermeasures for Internet of Things Networks: A Comprehensive Survey" *IEEE Access*, No. 9 (2021): 1-22.

23 European Union Agency for Cybersecurity (ENISA). *Threat Landscape for Artificial Intelligence* (Heraklion, Greece: ENISA, 2020).

24 Miles Brundage, Shahar Avin, Jack Clark, et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (Cambridge: 2018).

resilience, which is the ability of systems and societies to resist manipulative exploitation and the distortion of perception, is a fundamental issue to be considered in the field of cybersecurity.[25]

d. Competence and Responsible Use of AI

Adequate competence in using artificial intelligence responsibly is very important, so that risks are avoided, and if they arise, their effects are quickly and effectively removed, while future risks are prevented. Artificial intelligence competence is also highlighted in Article 4 of the Artificial Intelligence Act, according to which suppliers and users of AI systems are required to take measures to ensure, to the greatest extent possible, an adequate level of AI competence among their personnel and other persons who are involved in the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training, the context in which the AI systems are to be used, and the persons or groups against whom the AI systems are to be used. The requirement to take into account the technical knowledge, experience, education, and training, as well as the context in which AI systems are to be used allows the obligation to ensure an adequate competence level of its personnel to be applied proportionately to suppliers and users of AI systems. Such competence should specifically cover how AI systems function, their potential failure modes, the particular types of vulnerabilities they introduce, and the approaches required to secure them against malicious attacks, which has been referred to as "artificial intelligence literacy."[26]

The principle expressed in Article 4 of the AI Act should be taken into account in the application of any type of obligation under this regulation by suppliers and entities using AI systems. This is a meta-norm that is an implementation in the Artificial Intelligence Act of the principle of proportionality expressed in Article 5(4) of the Treaty on European Union.

The responsible use of AI systems involves the application of good practices, including an ethical approach, and avoiding undesirable or prohibited

25 *The Global Risks Report 2023*, https://www.weforum.org/publications/global-risks-report-2023/, [accessed: 22.06.2025); *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence* (Artificial Intelligence Act), COM(2021) 206 final, April 2021.

26 Karin Stolpe, Jonas Hallström, "Artificial intelligence literacy for technology education" *Computers and Education Open*, No. 6 (2024): 1-8.

practices. This ethical approach relies on fundamental principles that guide the development and deployment of AI: Transparency, Accountability, Fairness, and Privacy. Integrating these interconnected principles is crucial for building trustworthy AI systems. This benefits each AI stakeholder: decision-makers receive well-argued proposals, citizens have access to transparent information and are treated equally, and companies can be assured that their data is protected, helping to ensure economic privacy. Furthermore, adopting an ethical approach plays an important role in mitigating potential harms by ensuring protection is built in by design, as highlighted in research exploring AI ethics.[27]

e. Prohibited AI Practices under the AI Act

Article 5 of the Artificial Intelligence Act prohibits the following AI practices in relation to the marketing, commissioning, or use of an AI system:

1. that uses subliminal techniques beyond the person's conscious awareness, or intentional manipulative or deceptive techniques, the purpose or effect of which is to cause a significant change in the behavior of an individual or group of individuals by significantly impairing their ability to make informed decisions, thereby causing them to take a decision they would not otherwise take in a way that causes or is likely to cause them, another individual, or a group of individuals serious harm;
2. that takes advantage of the vulnerabilities of an individual or a specific group of persons due to their age, disability or particular social or economic situation, and the purpose or effect of the system is to make a significant change in the behavior of the individual or a person in that group, in a way that causes or is reasonably likely to cause serious harm to that individual or another person;
3. for the purpose of evaluating or classifying individuals or groups of individuals conducted over a specified period of time on the basis of their social behavior or known, inferred or predicted personal characteristics or personality traits, when it leads to one or both of the following:

27 Petar Radanliev, "AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development" *Applied Artificial Intelligence*, No. 1 (2025): 1-41.

   a. harming or disadvantaging some individuals or groups of individuals in social contexts that are unrelated to the contexts in which the data were originally generated or collected; or
   b. harming or disadvantaging some individuals or groups of individuals, which is unjustified or disproportionate to their social behavior or its severity;
4. to conduct risk assessments on individuals to evaluate or predict the risk of an individual committing a crime based solely on profiling or an assessment of the individual's personality traits and characteristics (however, this prohibition does not apply to AI systems used to support a human assessment of a person's involvement in criminal activity, which is already based on objective and verifiable facts directly related to criminal activity);
5. which creates or expands facial recognition databases by acquiring facial images from the Internet or CCTV footage;
6. for drawing inferences about an individual's emotions in the workplace or educational institutions, except in cases where an AI system is to be implemented or marketed for medical or security reasons;
7. that individually categorize individuals based on their biometric data to deduce or infer information about their race, political views, trade union membership, religious or philosophical beliefs, sexuality or sexual orientation (this prohibition does not include cases of labelling or filtering lawfully acquired biometric data sets, such as images based on biometric data, or categorization of biometric data in the field of law enforcement);
8. that uses real-time remote biometric identification systems in public spaces for law enforcement purposes, unless such use is absolutely necessary for one of the following purposes:
   c. the targeted search for specific victims of abduction, human trafficking, or sexual exploitation of human beings, as well as the search for missing persons;
   d. the prevention of a specific, substantial, and imminent threat to the life or physical safety of persons, or an actual and present or actual and foreseeable threat of a terrorist attack;
   e. the location or identification of a person suspected of having committed a crime for the purpose of investigating, prosecuting, or the enforcement of penalties for, specific crimes punishable by imprisonment in a Member State.

f. Human Supervision of High-Risk AI systems

In the case of artificial intelligence, attention must be paid to high-risk systems, which must be used responsibly and with as much care as possible, so as not to create risks with far-reaching consequences. As stipulated in Article 14 of the Artificial Intelligence Act, high-risk AI systems shall be designed and developed in such a way that they can be effectively supervised by humans during the period of their use. In the context of e-government, where AI systems may support decisions impacting citizens' rights, access to services, or legal standing, effective human supervision takes on particular importance. For example, when AI systems are used to support critical processes like the verification required for issuing an identity card or a driving license, human oversight is necessary to prevent errors that could lead to denial or identity issues.[28] Similarly, the use of AI to decide on granting social aid or initiating a tax audit requires high transparency and accountability in decision-making, particularly because such procedures can be highly intrusive for citizens.

Such supervision shall be aimed at preventing or minimizing risks to health, safety, or fundamental rights that may arise when such a system is used for its intended purpose or under conditions of reasonably foreseeable misuse, particularly when such risks persist despite the use of appropriate tools. It should be emphasized that supervision measures must be commensurate with the risk, level of autonomy, and context of use of the high-risk AI system in question. A human supervisor of such a system should be able to take such actions (including ignoring, disregarding, or reversing the operation of a high-risk AI system) that are effective, including the emergency shutdown of a high-risk AI system.[29] To effectively fulfill this role within a public service context, human supervisors need software that provides clear information to understand the AI's recommendations. This helps ensure that their final decision is based on transparent reasoning that can be clearly communicated to a citizen if needed. It also enables them to make a justified exception in specific circumstances, allowing for flexibility – such as considering a more understanding approach in

28 Benedikt Barthelmess, Jean Langlois, "Digital Identity: Legal and Economic Issues", https://ideas.repec.org/p/hal/wpaper/hal-04179570.html, [accessed: 23.06.2025].

29 Markus Mueck, Christophe Gaie, Dimitris Gkikas, "Introduction to the European Artificial Intelligence Act," [in:] *European Digital Regulations*, eds. Markus Mueck, Christophe Gaie (Cham: Springer, 2025): 63. (53-90)

debt recovery when a citizen is facing significant personal hardship, like the loss of a family member.[30]

g. The Role of Regulation: Balancing Innovation and Security

It has been pointed out that over-regulation can make it difficult to find appropriate applications for innovations, which can slow down their diffusion and use, which in turn can hinder the use of artificial intelligence systems for defense purposes, including in weapon systems, and lead to dependence on external suppliers and third-party weapon manufacturers. The war in Ukraine has shown how important innovation is on the modern battlefield, including the use of artificial intelligence systems.[31] However, there is a needed for regulation to curb the inappropriate use of artificial intelligence tools, specifically to prevent their use from leading to crimes against humanity. Moreover, the absence of regulation could lead to cybersecurity risks, and cyber attacks on artificial intelligence systems can harm not only human rights and freedoms, but also the foundations of the functioning of the state, including its economy or security, which must be constantly protected. Another example is the decisions made during a pandemic, when the constraints on citizens are high (confinement, mandatory vaccination, etc.) and should be carefully considered from a human perspective.[32]

The implementation of artificial intelligence tools must ensure an adequate level of protection of human freedoms and rights. Restrictions

30 Laszlo Horvath, Oliver James, Susan Banducci, Ana Beduschi, "Citizens' acceptance of artificial intelligence in public services: Evidence from a conjoint experiment about processing permit applications" *Government Information Quarterly*, No. 4, (2023): 1-18; Jean Langlois-Berthelot, Christophe Gaie, Jean-Fabrice Lebraty, "Epidemiology Inspired Cybersecurity Threats Forecasting Models Applied to e-Government", [in:] *Transforming Public Services – Combining Data and Algorithms to Fulfil Citizen's Expectations*, eds. Christophe Gaie, Mayuri Mehta, (Cham: Springer 2024): 151-174.

31 Paweł Pelc, "Akt w sprawie sztucznej inteligencji" *Myśl Strategiczna*, No. 1 (2025): 45.

32 Christophe Gaie, Markus Mueck, "10 – An artificial intelligence framework to ensure a trade-off between sanitary and economic perspectives during the COVID-19 pandemic", [in:] *Deep Learning for Medical Applications with Unique Data*, eds. Deepak Gupta, Utku Kose, Ashish Khanna, Valentina Emilia Balas, (Cambridge: Academic Press, 2022): 197-217.

on their use must not lead to a violation of human dignity. Interference that will not be proportional to the purpose of the restrictions may in some cases lead to a violation of human dignity, which is unacceptable, while each case of restriction of individual freedoms and rights should be dealt with separately, taking into account the circumstances of each case.[33] The Constitution of the Republic of Poland of 2 April 1997 (Journal of Laws of 1997, No. 78, item 483, as amended) expressly states in Article 31(3) that restrictions on the exercise of constitutional freedoms and rights may be imposed only by law and only if they are necessary for preserving the security or public order in a democratic state, for the protection of the environment, health, or public morals, or for the protection of the freedoms and rights of others. Such restrictions shall not infringe the very essence of the freedoms and rights.

3 Conclusion

Any new technology can not only be an enabler, but can also generate risks,[34] so it needs to be used responsibly to minimize them and effectively manage their impact. This is particularly important in the area of cybersecurity. New technologies, including artificial intelligence, have an important role to play in cyberspace, especially as the digital sector develops so rapidly. Artificial intelligence can protect cyberspace from attacks by predicting and neutralizing them, but if it is used as a tool for cyber attacks, it can cause great damage in cyberspace, so there must be

33 Małgorzata Czuryk, "Restrictions on the Exercising of Human and Civil Rights and Freedoms Due to Cybersecurity Issues" *Studia Iuridica Lublinensia*, No. 3 (2022): 32. For restrictions on human freedoms and rights, see also: Radosław Kostrubiec, *Dopuszczalne ograniczenia prawa do swobodnego, pokojowego zgromadzania się w systemie praw człowieka* (Zamość: Wydawnictwo Akademii Zamojskiej, 2024), 23; Małgorzata Czuryk, "Dopuszczalne różnicowanie sytuacji pracowników ze względu na religię, wyznanie lub światopogląd" *Studia z Prawa Wyznaniowego*, No. 27 (2024): 158, doi: 10.31743/spw.17518; Jarosław Kostrubiec, *Sztuczna inteligencja a prawa i wolności człowieka* (Warszawa: IWS, 2021), 21; Małgorzata Czuryk, "Activities of the Local Government During a State of Natural Disaster" *Studia Iuridica Lublinensia*, No. 4 (2021): 119-121. https://doi.org/10.17951/sil.2021.30.4.111-124.

34 Bogdan Grabowski, "Cyfrowe zagrożenia – zarys problemu" *Ius et Securitas*, No. 1 (2024): 103.

appropriate oversight mechanisms for its use. This supervision should be carried out by a human being.

ENISA notes that AI can be used to: create security mechanisms to detect, identify and mitigate breaches; exploit vulnerabilities in existing AI and non-AI tools and methods; and design a system to protect existing AI and non-AI tools and methods (protection created during system design).[35]

Artificial intelligence systems should be designed, modified, and implemented in such a way that they are cyber secure and that this appropriate level of cybersecurity is maintained throughout the life cycle of such systems. This will make it possible to prevent and debug such systems and to ensure appropriate protection when interacting with humans or other systems. Artificial intelligence systems that are placed on the market, put into service, or used should be adequately protected against unauthorized interference that could lead to their blocking, limiting their operation, altering their use, or manipulating the results they produce.

Integrating cognitive resilience into the regulation of cybersecurity and artificial intelligence will not only be a legal issue, but also a strategic imperative. Indeed, cognitive manipulation via AI has become a major vector of hybrid warfare, threatening social cohesion, national sovereignty, and the stability of democracies. The construction of a robust, coherent, and appropriate legal framework must therefore strengthen not only the technical security of systems, but also the resistance of societies to information manipulation. Furthermore, this legal development must be designed with international cooperation in mind, in order to avoid divergences that could weaken the collective response to hybrid threats. Finally, by ensuring that AI technologies always serve the rights, dignity, and freedoms of individuals, such regulations will strengthen trust in these systems, which are essential to national defense and security.

Societies that create and implement new solutions, especially in the field of AI, will be at a higher level of development than those that merely replicate them using AI tools. It is also important that new AI solutions always serve human beings, putting their dignity, freedoms, and rights first.[36]

35 Artificial Intelligence and Cybersecurity Research: ENISA Research and Innovation Bried (2023), 19.

36 *Polityka dla rozwoju sztucznej inteligencji w Polsce od roku 2020* (Warszawa: KPRM, 2020): 9.

Bibliography

Bierecki Dominik, Christophe Gaie, Mirosław Karpiuk, "Artificial Intelligence in e-Administration" *Prawo i Więź*, No. 1 (2025): 383-407.

Brundage Miles, Shahar Avin, Jack Clark, et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Cambridge: 2018.

Calo Ryan, "Artificial Intelligence Policy: A Primer and Roadmap" *U.C. Davis Law Review*, No. 2 (2018): 399-432.

Czuryk Małgorzata, "Restrictions on the Exercising of Human and Civil Rights and Freedoms Due to Cybersecurity Issues" *Studia Iuridica Lublinensia*, No. 3 (2022): 31-43.

Czuryk Małgorzata, "Activities of the Local Government During a State of Natural Disaster" *Studia Iuridica Lublinensia*, No. 4 (2021): 111-124, https://doi.org/10.17951/sil.2021.30.4.111-124.

Czuryk Małgorzata, "Cybersecurity and Protection of Critical Infrastructure" *Studia Iuridica Lublinensia*, No. 5 (2023): 43-52.

Czuryk Małgorzata, "Dopuszczalne różnicowanie sytuacji pracowników ze względu na religię, wyznanie lub światopogląd" *Studia z Prawa Wyznaniowego*, No. 27 (2024): 151-163, doi: 10.31743/spw.17518;

Gaie Christophe, Mirosław Karpiuk, Andrea Spaziani, "Cybersecurity in France, Poland and Italy" *Studia Iuridica Lublinensia*, No. 1 (2025): 73-95.

Gaie Christophe, Karpiuk Mirosław, Strizzolo Nicola, "Cybersecurity of Public Sector Institutions" *Prawo i Więź*, No. 6 (2024): 347-362

Gaie Christophe, Markus Mueck, "10 – An artificial intelligence framework to ensure a trade-off between sanitary and economic perspectives during the COVID-19 pandemic", [in:] *Deep Learning for Medical Applications with Unique Data*, eds. Deepak Gupta, Utku Kose, Ashish Khanna, Valentina Emilia Balas, (Cambridge: Academic Press, 2022): 197-217.

Gergelewicz Tomasz, "Bipolarity of Artificial Intelligence – Chances and Threats" *Ius et Securitas*, No. 2 (2024): 71-94.

Grabowski Bogdan, "Cyfrowe zagrożenia – zarys problemu" *Ius et Securitas*, No. 1 (2024): 95-105.

Horvath Laszlo, Oliver James, Susan Banducci, Ana Beduschi, "Citizens' acceptance of artificial intelligence in public services: Evidence from a conjoint experiment about processing permit applications" *Government Information Quarterly*, No. 4, (2023): 1-18.

Kaczmarek Krzysztof, "Finland in the light of cyber threats in the context of Russia's aggression against Ukraine" *Cybersecurity and Law*, No. 1 (2023): 204-214.

Kaczmarek Krzysztof, "Nordic countries in the face of digital threats" *Cybersecurity and Law*, No. 1 (2024): 151-161.

Kaczmarek Krzysztof, "Vulnerability to cyber threats: a qualitative analysis from societal and institutional perspectives" *Cybersecurity and Law*, No. 2 (2024): 106-116.

Kaczmarek Krzysztof, "Sztuczna inteligencja", [in:] *Leksykon cyberbezpieczeństwa*, ed. Katarzyna Chałubińska-Jentkiewicz (Warsaw: ASzWoj, 2024): 250-252.

Kaczmarek Krzysztof, "Wpływ zmian klimatycznych na bezpieczeństwo" *Journal of Modern Science*, No. 4 (2024): 410-430.

Kaczmarek Krzysztof, Mirosław Karpiuk, Claudio Melchior, "A Holistic Approach to Cybersecurity and Data Protection in the Age of Artificial Intelligence and Big Data" *Prawo i Więź*, No. 3 (2024): 103-121.

Karpiuk Mirosław, "Recognizing an Entity as an Operator of Essential Services and Providing Cybersecurity at the National Level" *Prawo i Więź*, No. 4 (2022): 166-179.

Karpiuk Mirosław, "The Legal Status of Digital Service Providers in the Sphere of Cybersecurity" *Studia Iuridica Lublinensia*, No. 2 (2023): 189-201.

Karpiuk Mirosław, "Glosa do wyroku Naczelnego Sądu Administracyjnego z dnia 12 lutego 2018 r. (II OSK 2524/17)" *Studia Iuridica Lublinensia*, No. 1 (2019): 185-194, doi: 10.17951/sil.2019.28.1.185-194.

Kostrubiec Jarosław, Mirosław Karpiuk, Dominik Tyrawa, "The status of municipal government in the sphere of ecological security" *Hungarian Journal of Legal Studies*, No. 2 (2024): 164-181. https://doi.org/10.1556/2052.2024.00510.

Kostrubiec Jarosław, *Sztuczna inteligencja a prawa i wolności człowieka*, Warszawa: IWS, 2021.

Kostrubiec Radosław, *Dopuszczalne ograniczenia prawa do swobodnego, pokojowego zgromadzania się w systemie praw człowieka*, Zamość: Wydawnictwo Akademii Zamojskiej, 2024.

Langlois-Berthelot Jean, "Evaluating and Insuring Cyber Risks within Organizations", Paris: EHESS, 2021.

Langlois-Berthelot Jean, Christophe Gaie, Jean-Fabrice Lebraty, "Epidemiology Inspired Cybersecurity Threats Forecasting Models Applied to e-Government", [in:] *Transforming Public Services – Combining Data and Algorithms to Fulfil Citizen's Expectations*, eds. Christophe Gaie, Mayuri Mehta, (Cham: Springer 2024): 151-174.

Mueck Markus, Christophe Gaie, Dimitris Gkikas, "Introduction to the European Artificial Intelligence Act", [in:] *European Digital Regulations*, eds. Markus Mueck, Christophe Gaie, (Cham: Springer, 2025): 53-90.

Pelc Paweł, "Akt w sprawie sztucznej inteligencji" *Myśl Strategiczna*, No. 1 (2025): 35-46.

Radanliev Petar, "AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development", *Applied Artificial Intelligence*, No. 1 (2025): 1-41.

Singla Alex, Alexander Sukharevsky, Lareina Yee, Michael Chui, Bryce Hall, "The state of AI: How organizations are rewiring to capture value", New York: McKinsey Global Publishing, 2025.

Stolpe Karin, Jonas Hallström, "Artificial intelligence literacy for technology education" *Computers and Education Open*, No. 6 (2024): 1-8.

Tkaczyk Edyta, "Bezpieczeństwo państwa w Konstytucji Rzeczypospolitej Polskiej. Refleksje nad dobrem chronionym" *Ius et Securitas*, No. 1 (2024): 39-54.

Tsagourias Elias, James Buchan, "Research Handbook on International Law and Cyberspace", Cheltenham: Edward Elgar Publishing, 2015.

Włodyka Ewa Maria, Krzysztof Kaczmarek, "Cyber Security of Electrical Grids – A Contribution to Research" *Cybersecurity and Law*, No. 2 (2024): 260-272.

Wojciechowski Tomasz, "Cyberbezpieczeństwo i dezinformacja we współczesnym świecie: strategie ochrony i zarządzania kryzysowego" *Ius et Securitas*, No. 1 (2024): 83-94.

Zaman Shakila, Khaled Alhazmi, Mohammed Aseeri, Muhammad Raisuddin Ahmed, Risala Tasin Khan, Shamim Kaiser, Mufti Mahmud, "Security Threats and Artificial Intelligence based Countermeasures for Internet of Things Networks: A Comprehensive Survey" *IEEE Access*, No. 9 (2021): 1-22.