ISSN 2949-2483

Volume 3 Number 3

JOURNAL OF DIGITAL TECHNOLOGIES AND LAW

2025

ELECTRONIC SCIENTIFIC AND PRACTICAL JOURNAL

JOURNAL OF DIGITAL TECHNOLOGIES AND LAW

Volume 3, No. 3
2025

DOI: 10.21202/2949-2483.2025.3

ELECTRONIC SCIENTIFIC AND PRACTICAL JOURNAL

eISSN 2949-2483

Published since 2023, frequency – 4 issues a year. DOI: 10.21202/2949-2483

Address: 42 Moskovskaya Str., 420111 Kazan, Russian Federation
Tel.: +7 (843) 231-92-90
Fax: +7 (843) 292-61-59
E-mail: lawjournal@ieml.ru
Website: https://www.lawjournal.digital
Telegram: https://t.me/JournalDTL_world
VKontakte: https://vk.com/JournalDTL
Yandex.Dzen: https://dzen.ru/JournalDTL
Odnoklassniki: https://ok.ru/JournalDTL

Founder and publisher of the Journal

Kazan Innovative University named after V. G. Timiryasov. Address: 42 Moskovskaya Str., 420111 Kazan, Republic of Tatarstan, Russian Federation. Tel.: +7 (843) 231-92-90. Fax: +7 (843) 292-61-59. E-mail: info@ieml.ru. Website: https://ieml.ru

Certificate on registering a mass medium: EL no. FS 77-84090 of 21.10.2022, issued by Roskomnadzor.

Territory of distribution: Russian Federation, foreign countries.

Date of signing the issue for publication: 2025, September 25. Hosted on the website https://www.lawjournal.digital: 2025, September 30.

JOURNAL OF DIGITAL TECHNOLOGIES AND LAW
Volume 3, No. 3
2025
DOI: 10.21202/2949-2483.2025.3
ELECTRONIC SCIENTIFIC AND PRACTICAL JOURNAL
eISSN 2949-2483

Content

Research article
UDC 34:004:341.4:004.8
EDN: https://elibrary.ru/gcvuaw
DOI: https://doi.org/10.21202/jdtl.2025.15

International Fundamentals of Legal Regulation of the Data Center Industry in the Arctic States and the Antarctic

Nataliya Igorevna Shumakova
South-Ural State University (National Research University), Chelyabinsk, Russia

Keywords

data processing centers,
digital technologies,
environmental law,
environmental security,
indigenous peoples,
international law,
law of indigenous peoples,
law,
the Antarctic,
the Arctic

Abstract

Objective: to critically assess the effectiveness of existing international legal norms under the new challenges of technological progress, related to the development of the data center industry in the Arctic states and the Antarctic.

Methods: the methodological basis of the research is a set of special and general methods of scientific cognition, including methods of comparative law, content analysis, deduction, induction, formal logical method and document analysis. The author turns to interdisciplinary approaches in order to objectively assess the environmental, social and legal risks arising from the data center industry growth in regions with increased climatic and social vulnerability.

Results: the article analyzed international legal acts regulating the functioning of data centers in polar regions. It identified the key risks and divided them into environmental (instability of local ecosystems, lack of adaptability to rapid changes, risk of losing biological diversity, and greenhouse gas emissions) and social (marginalization and violation of the rights of indigenous peoples, loss of traditional cultures and lifestyles, increased social tension). The author points out that new conflicts and challenges will inevitably emerge due to the insufficient effectiveness of national and international regulatory mechanisms. The states the need to create specialized international legal instruments taking into account the specifics of the environmental safety of the polar territories.

Scientific novelty: for the first time, the article provides a comprehensive analysis of the integral risks and drawbacks of the current international legislation on data center industry in the Arctic states and the Antarctic. The author provides a comparative analysis of the normative framework and shows the inconsistency between the "soft law" principles application on the polar regions and the fourth technological revolution. The author substantiates the requirement to create new certification and reporting procedures throughout the lifecycle of data centers, taking into account the legal and cultural context.

Practical significance: the results are focused on improving international and national policies in the sphere of regulating the data center industry and on developing certification and reporting standards that could be effective in the climatic, social and economic conditions of the Arctic states and the Antarctic. The research is aimed at minimizing the negative impact of anthropogenic factors and ensuring a balance between industrial development and the preservation of unique natural and cultural landscapes.

For citation

Shumakova, N. I. (2025). International Fundamentals of Legal Regulation of the Data Center Industry in the Arctic States and the Antarctic. *Journal of Digital Technologies and Law, 3*(3), 369–396. https://doi.org/10.21202/jdtl.2025.15

Contents

Introduction

The rapid pace of the ongoing fourth technological revolution is accompanied by the growth of new industries, one of which is the construction of data processing centers (further referred to as DPCs). Energy-intensive data centers that require maintaining a certain temperature are located up to the extreme points of the globe – the Arctic and the Antarctic. The Arctic states are attractive for locating data centers due to the relatively

low cost of electricity and to the climatic conditions that help reduce cooling costs. In Antarctica, the anthropogenic activity grows because we need to research its regions and popularize polar tourism. This also requires the creation of new data centers and the laying of deep-sea cables for operational communication and timely data transmission.

The vulnerability of the Arctic and Antarctic regions to anthropogenic emissions is a well-known fact, and an array of international legal acts are devoted to protecting their environment. Nevertheless, there is reason to believe that in the 21st century, not all states that have acceded to its conventions and declarations are sufficiently complying with their obligations. In particular, this is related to the implementation of the objectives of the EU Law on Critical Raw Materials, such as the independence of its member states from third countries in the provision of extractable resources[1]. There is an additional factor indicating a decrease in the effectiveness of international law on land use change in the Arctic countries: it is the violation of the rights of the indigenous peoples under the auspice of the critical need for mining required for the developing digital technologies and the transition to green energy. The vague formulations and criteria of current legal regulations make it possible to implement political decisions that go against the will of the indigenous peoples, even in those countries where they enjoy relative autonomy guaranteed at the constitutional level (Živojinović et al., 2024).

The above defines the article objective – to provide a critical assessment of the effectiveness of existing international legal norms under the new challenges of technological progress. This objective is achieved by identifying risk groups specific to the Arctic states and the Antarctic; analyzing general and special international legal acts; determining the role of these territories in ensuring the data centers development and operation; and identifying risks directly related to the growth of the data center industry in their regions. The main purpose of the study is to call for the development of special international measures that can offset the negative impact of new types of anthropogenic activity in the territories on which the climatic well-being of all humankind depends. The article calls for the development of uniform certification and reporting requirements for the entire lifespan of a data center, including accompanying industries. The author also calls for establishing liability violations, given the specifics of the environmental safety, cultural and social features of these territories.

1. The Arctic and Antarctic as special risk zones subject to international regulation

Under the global climatic changes, The Arctic and Antarctic are the zones of specific economic risks. Temperature rise results in rapid warming of the polar regions, provoking ice thinning, sea ice and eternal frost melting and causing negative consequences far

[1] Regulation (EU) 2024/1252 of the European Parliament and of the Council of 11 April 2024. https://clck.ru/3NNWKX

beyond their boundaries (Raimondi et al., 2024). The consequences include disturbance of climatic and geochemical cycles leading to not only the loss of biological diversity and animals' habitat, but also social-economic degradation of the arctic regions' population. Moreover, today the point of no return may have been reached, which means the humankind may face "a cascade of unfavorable consequences" for the whole planet[2]. According to the research under Horizon 2020 project held by University of Bergen, the Arctic is most acutely influenced by the global warming caused by human activity. In the Arctic Ocean, acidification due to greenhouse emissions is progressing 10 times, and warming – two times faster than anywhere in the world. This testifies to passing the point of no return as an actual fact[3]. For example, together with the increased carbon dioxide (CO_2) content in the atmosphere and temperature growth, seasonal fluctuations of partial carbon dioxide (pCO_2) grow, as well as the changes in water hydrogen index (pH). This increases the summer acidification of the ocean, which may reduce the endemic sea organisms' resistance to higher summer temperatures (flying snails, copepods, polar cod, etc.). These organisms are the key link between zooplankton and sea mammals, sea birds and other fishes (Orr et al., 2022). Increased CO_2 emissions is the primary cause of the water surface absorbing the anthropogenic carbon (C_{ant}), which, in turn, is the main factor of the open ocean acidification (Terhaar et al., 2020). At the same time, the share of oxygen (O_2) determines the safety of sea ecosystems, sensitive to the interconnected processes of warming, acidification, deoxygenation[4] (reduced oxygen level), decreased nutrients and primary production[5]. Scientists also point to the so called polar amplification – the proved fact that the Arctic warms faster than the rest of the globe. Recent research demonstrate that from 1979 to 2021 this process was almost four times faster than anywhere else. This indicates that earlier forecasts of the global warming speed (two times faster than in other regions) were not accurate, insufficiently evaluating the situation on the recent 43 years. Since the start of satellite observations, certain regions in the Eurasian sector of the Arctic Ocean warmed up to seven times faster than the globe, while the overall temperature growth in the Arctic was almost four times faster than in other territories during the same period (Rantanen et al., 2022). The factors of climate change in the Arctic are also globalization and digitalization. The territory contains deposits of mineral resources, which are necessary for the "green shift" – the transition to renewable energy sources. As the European Union strives for zero emissions of carbon dioxide by 2050, their excavation

2 EU. Understanding impacts of climate change on Earth's vulnerable polar regions. https://clck.ru/3NNWzc

3 EU. (2020). Our common future ocean in the Earth system. https://clck.ru/3NNX3x

4 FAQ: Ocean Deoxygenation. Scripps Institution of Oceanography. https://clck.ru/3NNX5m

5 Ibid.

(for production of solar panels, wind turbines, automobile accumulators, etc.) continues to grow. This entails irreversible impacts on the environment and landscape and causes alarm (and sometimes social tension) among the local population (Živojinović et al., 2024). Positive aspects of the industrial development of these territories include infrastructure development, new jobs, increased income of the local population and higher tax revenues. However, there are also risks of disappearance of traditional industries and other elements of national cultures, which leads to the loss of cultural identity (Živojinović et al., 2024). Indeed, the development of new types of land use, including excavation of raw materials for the "green shift" and other purposes, electric energy production, construction of roads and other infrastructure are the drivers of negative changes in traditional deer breeding in Fennoscandia (according to 60 % of respondents in a poll held within an Artic Hub project) (Turunen et al., 2024). Given the current geopolitical situation, mining causes more and more disputes – on the one hand, the EU strives for self-provision with all necessary mineral resources. On the other hand, fundamental rights of the residents, as well as the research ethics, may be violated. This indicates further exacerbation of the "war for resources" not only among countries but also between states and population of the Arctic. The latter is more and more unsatisfied by turning their lands into "Eldorado for large companies" and by perceived lack of state defense of their legal rights and interests (Suopajärvi et al., 2024). It should be understood that the mentioned phenomena are not a problem of the future generations. In particular, meteorological observations in the Icelandic Westfjords show that the climate started to change quicker in the few recent years. Before the end of the century, this arctic state will face the corresponding changes of the physical and anthropogenic environment, and the speed of these changes depend on reduction of anthropogenic emissions (Bannan et al., 2022). Scientists more and more often mark the correlation between the Arctic temperature rise and increased appeals for psychological help among the local population. The latter develop "ecological anxiety", exacerbation of solastalgia and psychological states such as "climatic and environmental grief". These are reactions to the global negative changes in the environment, including those associated with changes in land use (construction of mines, railroads, wind power stations, etc.), and the lack of power in making ecological decisions when their rights are restricted by policies (Markkula et al., 2024). This is also about safety – the eternal frost is a reservoir of biological, chemical and radioactive materials; hence, the ongoing melting leads to wakening of the ancient, often unknown primitive organisms. This increases the risk of biological danger, including beyond the Arctic (Ali et al., 2024). Besides, the melting of the eternal frost, including the underwater one in the sea, may

cause emissions of another greenhouse gas – methane (CH4), contained in the natural sources of the region (wetlands, fresh water systems, gas hydrates, etc.) and produced as the soil humidity rises (Parmentier et al., 2024). Together with the increased frequency of extreme weather phenomena, these negative factors have already led to mass loss of flora and fauna, caused changes in migration of birds and fishes, led to disturbance of coastal social-economic systems and reduction (and sometimes complete elimination) of fish and other sea resources (Pecuchet et al., 2025). In the context of health of the Arctic population and wild nature, it is necessary to mention the ongoing pollution of the region with mercury (Hg)[6].

The term "the point of no return" started being used in relation to the Antarctic too. This is the ice shield of the planet, containing over 60 % of the world stock of fresh water. Climate changes lead to ice melting, ice shelf disintegration and, as a consequence, to increased ocean level[7]. Splitting of ice masses leads to water mixing and redistribution of warmth in the ocean, disturbing delivery of nutrients into the euphotic zone. This violates the stability of the ocean's upper layer, hence, the availability of light necessary for plankton, influencing the CO2 absorption (Meredith et al., 2022). Experiments show that with the temperature growth, the speed of ice melting increases, which will lead to the fast desalination of the continental shelf (Mathiot & Jourdain, 2023). In Antarctic, these phenomena are not due to greenhouse gases emissions on its territory, but are the result of global changes in the atmosphere and ocean. They go with the same speed as, for example, in Greenland, where such connection is established. This does not make the climatic forecast more optimistic, however. Modeling demonstrates that, with the current policy preserved, the sea level will rise not less than by 42 cm (Edwards et al., 2021). The negative influence on the continent's climate is produced also by extreme temperature phenomena (heat waves and sea heat waves), which may provoke a cascade of extreme events, such as a record temperature rise in the East Antarctic caused by an atmospheric river or a complete disintegration of a shelf Conger Glacier in 2022 (Siegert et al., 2023). The Antarctic lack local population, but its territories are influenced by anthropogenic activity related to science and tourism (appearance of non-local flora and fauna, mobilization of pollutants from waste deposits due to ice melting, etc.) (Hughes et al., 2021). Researchers call for paying attention to the lack of the necessary norms and criteria to assess pollution caused by increased human activity (Bargagli & Rota, 2024).

6 Why is mercury a concern in the arctic? AMAP. https://clck.ru/3NNXMK

7 EU. (2020). Identifying ice loss 'tipping points' in Antarctica. https://clck.ru/3NNXP2

1.1. General international legal framework for ensuring environmental safety and respect for the rights of indigenous peoples

The United Nations (further referred to as the UN) 2030 Agenda for Sustainable Development A/RES/70/1, within the list of Sustainable Development Goals (further referred to as the SDGs), highlights the need to take urgent measures to combat climate change and its consequences. This cannot be achieved without fulfilling the commitments made by developed countries – parties to the UN Framework Convention on Climate Change (further referred to as the UNFCCC), to accomplish joint mobilization (Goal #13)[8], which underlies all environmental standards. These include: reporting on anthropogenic emissions; development of climatic change mitigation programs; promotion and cooperation in the field of technologies for minimizing anthropogenic emissions; cooperation in taking preparatory measures to adapt to the climatic change; taking into account the considerations related to climatic change in the implementation of the relevant social, economic and environmental policies and measures; promotion and cooperation in the comprehensive, open and prompt exchange of scientific, technological, technical, socio-economic and legal information related to the climatic system and climatic change; implementation of national policies and appropriate measures by developed countries to mitigate the effects of climatic change by limiting their anthropogenic greenhouse gas emissions and protecting and improving the quality of their sinks and reservoirs of greenhouse gases; provision of new and additional financial resources to cover all agreed costs caused by the Convention fulfillment by developing countries and other obligations provided for in Art. 4[9]. A number of agreements have also been signed within the UNFCCC framework, including:

1. The Kyoto Protocol (1998). Its key provisions are commitments to reduce greenhouse gas emissions into the atmosphere, including CH4 and CO2, to establish transparency in the field of anthropogenic emissions and responsibility, given the features of the parties' economic development[10].

2. The Bali Action Plan (2007). It recognizes the fact of global warming to be proved and calls for the development and strengthening of measures to combat it, including technological, financial and political ones[11].

3. The Copenhagen Agreement (2009). It considers global critical climate change as one of the main problems of humanity, which has a scientific basis and requires urgent solutions. To this end, we must reduce anthropogenic emissions, intensify actions and

8 United Nations. (2015, October 21). Resolution adopted by the General Assembly on 25 September 2015. https://clck.ru/3NNZvD

9 UN Framework Convention on Climate Change. Adopted on May 9, 1992. https://clck.ru/3NNZxc

10 UNO. (1998). The Kyoto Protocol. https://clck.ru/3NNa4f

11 United Nations. (2007, December 14). FCCC/CP/2007/L.7/Rev.1. https://clck.ru/3NNa6R

international cooperation in reducing global and national greenhouse gas emissions, with more developed countries supporting less developed ones. It is also important to combat cutting and degradation of forests, in order to maintain greenhouse gas uptake, preserve biodiversity and living conditions of indigenous peoples[12].

4. The Cancun Agreements (2010). They are also aimed at increasing transparency in the area of annual greenhouse gas emissions, as well as reducing them depending on the economic development of a country[13].

5. The Durban Platform (2011). It provides for the continuation of the Kyoto Protocol and establishes the structure of the Green Climate Fund, whose task is to support adaptation to climatic change in less economically developed countries[14].

6. The Paris Agreement (2015). It once again raises issues of accountability and transparency, calling for real action to fulfill the commitments made by the UNFCCC member states, to take action under the Kyoto Protocol, to keep global average temperature rise well below 2 °C above pre–industrial level, and to strive to limit temperature rise up to 1.5 °C above pre-industrial level, recognizing the climatic change in order to reduce its risks and effects[15].

The general international framework for the protection of environmental safety and the rights of indigenous peoples in the polar regions also includes:

1. The Convention on the Continental Shelf (1958). It obliges coastal states to take safety measures to protect marine living resources from harmful effects[16].

2. The International Covenant on Economic, Social and Cultural Rights (1966). It prohibits the deprivation of peoples of their means of subsistence[17].

3. The Convention for the Prevention of Marine Pollution by Dumping of Wastes and Other Materials (1975). It is aimed at effective control of marine pollution and imposes obligations on member states to control dumping[18].

4. The Convention on Long-range Transboundary Air Pollution (1979). It is aimed at protecting humans and the environment from air pollution[19].

---

12 United Nations. (2009, December 18). FCCC/CP/2009/L.7. https://clck.ru/3NNa8z

13 United Nations. (2011, March 15). FCCC/CP/2010/7/Add.1. https://clck.ru/3NNaGB

14 The UNO Durban Platform. (2013, July). https://clck.ru/3NNaHD

15 United Nations. (2015, December 12). FCCC/CP/2015/L.9/Rev.1. https://clck.ru/3NNaJE

16 The Convention on the Continental Shelf. (1958). https://clck.ru/3NNaL3

17 UNO. (1966, December 16). The International Covenant on Economic, Social and Cultural Rights. https://clck.ru/3NNaM5

18 UNO. (1975). The Convention for the Prevention of Marine Pollution by Dumping of Wastes and Other Materials. https://clck.ru/3NNaNq

19 UNO. (1979). The Convention on Long-range Transboundary Air Pollution. https://clck.ru/3NNaS5

5. The United Nations Convention on the Law of the Sea (1982). Among other things, it stipulates the right of coastal states to take legislative measures to preserve the environment and prevent its pollution[20].

6. The Vienna Convention for the Protection of the Ozone Layer (1985). Its main purpose is to protect human health and the environment from the effects of changes in the ozone layer caused by anthropogenic activity[21].

7. The Convention on Indigenous and Tribal Peoples in Independent Countries (1989). It obliges the participating countries to promote the full realization of the social, economic and cultural rights of indigenous peoples while respecting their social and cultural identity, and, if necessary, to take special measures to protect the peoples concerned, their institutions, labor, culture, and the environment[22].

8. The Convention on Environmental Impact Assessment in a Transboundary Context (1991). It requires parties to take all possible measures to prevent and control significant harmful transboundary impacts resulting from planned activities and control over them, such as environmental impact assessment[23].

9. The Convention on Biological Diversity (1992). It is aimed at countering the loss of the planet's biological diversity, recognizing the dependence of indigenous peoples on biological resources and the need to use them mutually on an equitable basis. It also obliges the participating countries to create specially protected areas, take measures for rehabilitation and restoration of endangered species and interact with indigenous peoples for these purposes, based on the principles of respect, preservation and support of their knowledge and traditions[24].

10. The United Nations Declaration on the Rights of Indigenous Peoples (2007). It prohibits any actions aimed at depriving indigenous peoples of their lands, territories and resources, including actions that may result in the destruction of culture and identity. At the same time, it gives indigenous peoples the right to participate in decision-making on issues that would affect their rights[25].

11. The Stockholm Convention on Persistent Organic Pollutants (2011). It recognizes the particular vulnerability of ecosystems and communities in the Arctic due to the bio-amplification of exposure to persistent organic pollutants and contamination

---

20 UNO. (1994). The United Nations Convention on the Law of the Sea. https://clck.ru/3NNaVL

21 UNO. (1985). The Vienna Convention for the Protection of the Ozone Layer. https://clck.ru/3NNaa2

22 UNO. (1989). The Convention on Indigenous and Tribal Peoples in Independent Countries. https://clck.ru/3NNabb

23 UNO. (1991). The Convention on Environmental Impact Assessment in a Transboundary Context. https://clck.ru/3NNadD

24 UNO. (1992). The Convention on Biological Diversity. https://clck.ru/3NNmeG

25 UNO. (2008, March 17). The United Nations Declaration on the Rights of Indigenous Peoples. https://clck.ru/3NNccN

of traditional food products used by indigenous peoples. It also requires taking measures to reduce or eliminate emissions from anthropogenic activity[26].

12. The Minamata Convention on Mercury (2013). It aims to protect human health and the environment from anthropogenic emissions and releases of mercury and its compounds[27].

The list also includes international legal tools adopted by the International Maritime Organization (IMO), such as the International Convention for the Prevention of Pollution from Ships MARPOL 73/78 (aimed at combating ocean pollution), the International Safety Code for Ships Using Gases or Other Fuels with a Low Flash Point (MGT Code 2017), the International Convention on the Control and Management of Marine Ballast Water and Sediments (2004), etc. IMO is also developing medium-term measures to reduce greenhouse gas emissions from ships and the use of hydrogen and ammonia as marine fuels[28].

1.2. Special international legal framework for ensuring the safety and rights of indigenous peoples in the Arctic and Antarctic

Given the strategic, social, economic and climatic features of the polar regions, an extensive multi-level international legal framework has been formed to ensure security on their territory. Besides general international legal framework for ensuring environmental safety and protecting the rights and legitimate interests of indigenous peoples, a whole range of international acts has been developed. They are aimed at modifying national legislation, developing comprehensive international cooperation, and increasing transparency in the field of anthropogenic emissions in the Arctic and Antarctic. For this study, the main special international legal acts that ensure security in these regions are:

1. The Antarctic Treaty (1959). It prohibits nuclear explosions in Antarctica and the disposal of radioactive materials in this area, establishes the principles of transparency of work and scientific research, establishes control over expeditions and stations, and calls for cooperation in the development of measures for the protection and conservation of living resources in Antarctica[29].

2. The Protocol on Environmental Protection to the Antarctic Treaty (1991). Its parties assume responsibility for the "comprehensive protection of the Antarctic environment and its dependent and associated ecosystems"[30]. In this connection, it is recognized necessary to limit the negative impacts on the Antarctic environment and its dependent

26 The Stockholm Convention on Persistent Organic Pollutants. (2001). https://clck.ru/3NNcdR

27 UNO. (2013). The Minamata Convention on Mercury. https://clck.ru/3NNchn

28 International Maritime Organization. Official Internet resource of the Ministry of Transportation of the Russian Federation. https://clck.ru/3NNcjJ

29 The Antarctic Treaty. (1959). https://clck.ru/3NNck5

30 The Protocol on Environmental Protection to the Antarctic Treaty (1991). https://clck.ru/3QHnur

and associated ecosystems (negative effects on climate, weather, water and air quality, ice and marine environments, flora and fauna)[31].

3. The Declaration of the Protection of the Arctic Environment (1991). It is aimed at preserving the environment and natural resources, monitoring its condition and reducing pollution, as well as harmonizing environmental principles with the needs of the indigenous population[32].

4. The Agreement on Consent and Cooperation between the Russian Federation and Canada (1992). It emphasizes the role of the participating countries in environmental conservation and aims, among other things, at strengthening their cooperation in the Arctic, which is considered as a priority area of Russian-Canadian relations. The Agreement also provides for constant interaction with the indigenous peoples of the northern regions[33].

5. The Russian-Swedish Declaration (1993). It establishes international cooperation between the Russian Federation and Sweden, implying "socially and environmentally oriented policies, economic liberalization, freedom of trade and entrepreneurship within the framework of a civilized attitude to the environment and the rational use of natural resources"[34].

6. The First Kirkenes Declaration (1993). It set the basic principles of cooperation in the Barents/Euro-Arctic region and established the Council of Barents/Euro-Arctic region. The declaration emphasized the importance of scientific and technological cooperation in the region, of the development of cultural relations and support for indigenous peoples (Nenets and Sami). To this end, it was proposed to create a special working group[35].

7. The Declaration on the Foundations of Relations between the Russian Federation and the Kingdom of Norway (1996). It was aimed at fruitful cooperation in the Barents/Euro-Arctic region, strengthening respect for human rights and fundamental freedoms, including the rights of national minorities. It also proposed intensifying work on global environmental issues and addressing issues related to the internal relationship between energy, environmental protection, and economic development[36].

8. The Iqaluit Declaration (1998). The participating countries committed themselves to improving the well-being of Arctic residents, as well as taking measures to protect and improve the environment, economy, culture and health of indigenous peoples and other peoples living in the region[37].

---

31 The Protocol on Environmental Protection. (1991). https://clck.ru/3NNcm2

32 The Declaration of the Protection of the Arctic Environment. https://clck.ru/3NNcpo

33 The Agreement on Consent and Cooperation between the Russian Federation and Canada. (1993). https://clck.ru/3NNcs3

34 The Russian-Swedish Declaration. (1993). https://clck.ru/3NNct2

35 Declaration Cooperation in the Barents Euro-Arctic Region. (1993, January 11). https://clck.ru/3NNcut

36 Declaration on the Foundations of Relations between the Russian Federation and the Kingdom of Norway. (1996). https://clck.ru/3NNcwu

37 The Iqaluit Declaration. (1998). https://clck.ru/3NNcyX

9. The Inuvik Declaration on Arctic Climate Change and Global Action (2005). It called for uniting the humanity in order to drastically reduce anthropogenic emissions to prevent further critical climate change, where the Arctic is one of the key components of the planet's climate well-being. The Declaration also recognizes that current climate change poses an existential threat to the safety of Arctic natives[38].

10. The Ilulissat Declaration (2008). It confirmed the unique status of Denmark, Canada, Russia, the USA, Norway and Canada as states capable of finding a solution to the climate crisis in the Arctic. This task does not require a special legal regime for the Arctic Ocean – the existing norms of international law and national regulations of the participating countries are sufficient, but only with their continued comprehensive cooperation[39].

11. The Declaration of the Circumpolar Inuits on Sovereignty in the Arctic (2009). On behalf of the Inuit living in Greenland, Canada, the USA and Russia, it reminded states that "in pursuit of economic opportunities in the Arctic, which continues to warm up", it is necessary, among other things, to strive for environmental sustainability, to prevent the harmful exploitation of resources and the marginalization of the indigenous population[40].

12. The Nuuk Declaration on Environment and Development in the Arctic (2010). It recalled that the Inuit are a single people living in four different countries, but united by a respectful attitude towards the shared culture, resources and "life itself" with other peoples. It recognized the fact that the rights of indigenous peoples, including Inuit, along with fundamental human rights, have not yet been fully implemented. The document indicated fragility of the Arctic environment under the increasing land and water resources development. It called for the exchange of knowledge with the indigenous population and their more active participation in the protection of the territories where they live[41].

13. The Declaration following the meeting of the Heads of Government of the BEAC member countries (2013). It confirmed the commitment of the participating countries to the principles of the First Kirkenes Declaration and focused on environmental protection and the protection of the rights of the indigenous population. The latter include the right to participate in decision-making on issues affecting their rights, as well as the rights of indigenous peoples to preserve the traditional way of life, including hunting, fishing and reindeer husbandry[42].

14. The International Code for Ships Operating in Polar Waters (Polar Code 2014). It was adopted in order to improve the safety of ship operation and limit its impact on

---

38 The Inuvik Declaration. (2008, December 5). https://clck.ru/3NNd5z

39 The Ilulissat Declaration. (2008, May 27–29). https://goo.su/vuqV

40 A Circumpolar Inuit Declaration on Sovereignty in the Arctic. (2009). https://clck.ru/3NNdEn

41 Nuuk Declaration. (2010). https://clck.ru/3NNdFu

42 The Declaration following the meeting of the Heads of Government of the BEAC member countries. (2013). https://clck.ru/3NNdGr

humans and the environment. In its Preamble, it recognizes that "communities of Arctic coastal peoples may be, and polar ecosystems are vulnerable to human activities such as shipping", and in Part II-A, it establishes pollution prevention measures (operational and structural requirements)[43].

15. The Reykjavik Declaration (2021). It recognizes the inextricable link between human, animal and environmental health and calls for further development of cooperation in the field of safety and health of Arctic communities and the social well-being of Arctic residents. It also calls for continued research on new, emerging and regulated pollutants and for strengthening measures to implement commitments, related to mercury (Hg) and carbon trioxide (CO3) pollution[44].

16. The Ilulissat Declaration (2022). It confirmed the status and goals of the Inuit Circumpolar Council, recognizing the entry of the Inuit and the rest of humanity into an era of environmental and global insecurity and condemning threats to food security, changes in wildlife, as well as environmental and industrial impacts on the territories of Inuit communities[45].

17.The Helsinki Declaration on Climate Change in Antarctica (2023). It recognized that if CO_2 emissions remain at current levels, the atmosphere and oceans will continue to heat up and the oceans – to acidify. The Declaration once again confirmed that mining for any purpose other than scientific research is prohibited on the continent, and stated the need to jointly study the impact of global climate change on Antarctica, as well as the role of Antarctica and the Southern Ocean in regulating the global climate and future rise of sea level[46].

The central organization for cooperation in the Arctic is the Arctic Council (further referred to as the AC), established in 1996 by the Arctic states (Russia, Denmark, Norway, Sweden, Finland and Canada). Its goals are environmental protection and sustainable development of the region[47]. The AC, as an intergovernmental forum, includes six permanent participants representing the interests of the indigenous peoples of the Arctic: the Aleutian International Association, the Athabaskan Arctic Council, the International Gwich'in Council, the Inuit Circumpolar Council, the Sami Union, and the Russian Association of Indigenous Peoples of the North. The AC also has special working groups, such as the AMAR International Charitable Foundation, which assesses the state of the Arctic environment, and PAME, conducting policy research and developing measures to protect the marine environment[48]. In addition to the AC, the operating organizations

43 The International Code for Ships Operating in Polar Waters (Polar Code). (2014). https://clck.ru/3NNdQw

44 The Reykjavik Declaration. (2021). https://clck.ru/3NNdST

45 Inuit Circumpolar Council. The Ilulissat Declaration of 2022. (2022). https://clck.ru/3NNdV3

46 The Helsinki Declaration on Climate Change in Antarctica. (2023). https://clck.ru/3NNdXN

47 Arctic Council. https://clck.ru/3NNdZC

48 Arctic Fund. AMAP, AEPS, CAFF, PAME International programs. https://clck.ru/3NNdaX

are: 1) The International Barents Secretariat, established under the Agreement between the governments of Finland, Norway, the Russian Federation and Sweden to develop cooperation in the Barents/Euro-Arctic region[49]; 2) Barents/Euro-Arctic Region Council (BEAC), established in 1998 to maintain cooperation in the field of environmental protection and improve the situation of the indigenous population of the North[50]; 3) the Barents Regional Council (BRC), which is an independent multilateral cooperation body of 13 administrative-territorial entities of the Barents region, comprising their leaders and representatives of indigenous peoples (Sami, Veps and Nenets) [51]; 4) an independent Working group on indigenous peoples with the status of a permanent advisory body to the BEAC and the BRC[52]. To date, the Russian Federation has denounced the Agreements between the governments of Finland, Norway, the Russian Federation and Sweden on the establishment of the International Barents Secretariat for the Development of Cooperation in the Barents/Euro-Arctic region (Decree of the Russian Government No. 921-p dated 04/16/2025[53]) and was excluded from the BEAC, which Finland also left[54].

In Antarctica, the role of a special international body is performed by the Antarctic Treaty Secretariat, which has established an Environmental Protection Committee. The functions of the latter are to present considerations and formulate recommendations to the parties in connection with the implementation of the Protocol on Environmental Protection, including the operation of its Annexes, for consideration at Antarctic Treaty Consultative Meetings (Article 11 of the Protocol on Environmental Protection)[55].

2. The role of the Arctic and Antarctic in ensuring the data centers development and operation

Over the past few years, a number of successful projects have been implemented in the Arctic to locate data processing centers (further referred to as DPCs). The latter appear to be a new type of critical infrastructure. The main reason for their location was the abundant reserves of electricity, including green one, and its low price, coupled with a cold climate,

49 Agreement between the governments of Finland, Norway, the Russian Federation and Sweden. (2007). https://clck.ru/3NNdhK

50 Barents/Euro-Arctic Region Council (BEAC). (2021, July 1). https://clck.ru/3NNdic

51 Barents Regional Council. (2021, July 2). https://clck.ru/3NNdjX

52 Barents/Euro-Arctic Region Council (BEAC). (2021, July 1). https://clck.ru/3NNdmL

53 Decree of the Russian Government No. 921-r of 16.04.2025. (2025). https://clck.ru/3NNdnn

54 Cabinet of Ministers denounced the agreement on cooperation in Barents region. (2025, April 18). TASS. https://clck.ru/3NNdob

55 Protocol on Environmental Protection to the Antarctic Treaty of October 4, 1991. (1998). https://clck.ru/3NNdpS

fiber-optic connectivity, and reasonable land prices (Saunavaara & Laine, 2021). For example, Iceland is positioning itself as an ideal location for DPC construction due to its cheap green energy and strives to become the largest data center operator in the Nordic countries[56]. Indeed, assessing its prospects for the DPC industry development, the KPMG international audit and consulting corporation highlighted such advantages as the possibility of using natural cooling (ventilation systems without additional coolers) and fairly low electricity prices (while energy consumption in Iceland increased by 75 % on average over a five-year period in 2018)[57]. The growth of requests for the construction and expansion of existing DPCs continues along with an increase in projects using AI technologies and the amount of data that needs to be processed. In particular, Verne in Iceland owns a DPC using hydro and geothermal energy. According to the company, it "does not cause any damage to the planet", but uses liquid cooling technologies[58]. It is noteworthy that Verne does not specify on its official page which type of liquid cooling is used – indirect water, immersion using a special mineral oil or some other technology. The Kolos data center, which was supposed to be located in the Arctic, in Norwegian Ballangen, also deserves attention. It was planned but never built due to changes in Norwegian legislation regarding miners. Claimed to be the largest in the world, it would be located on 600,000 square meters and use climate cooling, operating on hydroenergy[59]. In an effort to promote transparency and increase trust, the Norwegian Association of the Data Center Industry (Norsk Datasenterindustri) presented a report on DPCs for 2023–2024. It once again emphasized that the growth of this industry is due to digitalization and the widespread introduction of artificial intelligence, which causes additional CO2 emissions (in 2024, these emissions were 12 times higher in the UK and Germany than in Norway). The report confirmed the importance of data centers for the country's economy and recalled that this industry was brought under control through amendments to the Law "On Electronic Communications"[60]. Finland, which hosted the Meta[61] data center, is not far behind. In particular, one of their data centers, which processes huge amounts of data, is located in Lulea, Lapland. According to the company, the goal of zero CO_2 emissions was achieved there by using "clean and renewable energy" and a strategy of adding renewable energy sources to the local

56 Moss, S. (2024, 26 March). Iceland's AI moment. https://clck.ru/3NNeJD

57 The Icelandic Data Center Industry. (2018, March). https://clck.ru/3NNeLi

58 High-performance computing in Iceland. Verne. https://clck.ru/3NNeNQ

59 Kolos Data Center. https://clck.ru/3NNoHk

60 The Data Center Industry in Norway 2023–2024. (2024). Norwegian Data Center Industry. https://clck.ru/3NNeX4

61 The organization is recognized as extremist, its functioning is prohibited in the territory of the Russian Federation.

network. Moreover, the company claimed that it would "restore more water than consume" by 2030[62].

In Russia, both non-profit and commercial data centers also continue to grow. Technological modules with a total capacity of 4,000 devices with a capacity of 16 MWatt will soon be placed directly above the Arctic Circle on a land plot of 15,000 square meters. These are part of a data center owned by Intelion Sever, a resident of the Russian Arctic zone[63]. To avoid the impact of sanctions, the state-owned Rosatom Corporation is also planning to launch a data center in the Murmansk oblast. It will be based on the Kola Nuclear Power Plant, where about 20–25 % of electricity remains unclaimed, while the cold climate makes it possible not to install cooling systems, also reducing electricity consumption[64]. A Russian hosting provider RUVDS went even further and in 2024 launched a modular data center on a drifting ice floe, directly next to the North Pole. It was powered by diesel generators, as part of an experiment that was completed within a month due to the appearance of a crack on the ice floe[65]. RUVDS has planned a similar project for 2025, but this time the data center will be located at the South Pole[66].

Chile, which is also striving to become a country that attracts data center operators, plans to extend underwater cables to the last of the continents deprived of them, to the Antarctic. The Antarctic SMART Cable, with almost unlimited bandwidth, is to connect Antarctica's largest research center to either the American McMurdo Station, or New Zealand's Invercargill, or Australia's Sydney. This would improve current and future Antarctic research and create the opportunity for stable interaction for scientists and staff[67].

3. Risks associated with the growth of the data center industry and attempts to mitigate them

The first group of risks associated with the construction of data centers in the Arctic states and the Antarctic certainly comprises the risks of their negative impact on the environment. The Norwegian research center SINTEF Energi AS draws attention to the fact that when electric power is used to cool the data center (air cooling of data processors), an excessive amount of heat is generated at a temperature of 40–50 °C. This temperature

62 Meta's* Luleå Data Centre. https://clck.ru/3NNedg (* The organization is recognized as extremist, its functioning is prohibited in the territory of the Russian Federation)

63 First commercial DPC in Murmanskoblast to be launched by a Russian Arctic zone resident with state support. (2023, July 13). Corporation for the Far East and Arctic development. https://clck.ru/3NNeje

64 "Arktika" DPC at Kola nuclear power station will be built solely with Russian equipment. (2022, July 18). TASS. https://clck.ru/3NNeoC

65 First DPC in the Arctic! RuVDS. https://clck.ru/3NNepL

66 RUVDS will test server equipment in the Antarctic. (2024, October 16). RuVDS. https://clck.ru/3NNeqT

67 Winston Qiu. (2024, December 14). US NSF Requests for Information on Antarctic SMART Cable. Submarine cable networks. https://clck.ru/3NNevG

increases up to 60-80 °C if more efficient cooling systems are used (liquid or two-phase, in which the liquid refrigerant evaporates in a cold plate heat exchanger). In both cases, this excess heat is usually not used in any way; moreover, CO2 emissions from data centers already account for at least 2 % of the global total, which is equivalent to damage from the aviation industry[68]. Excess heat, if it has a sufficiently high temperature, can be used to heat buildings and for other industrial and household purposes, as is done, for example, in Sweden (Yuan et al., 2023). However, today excess heat is most often released into the atmosphere, which, for example, is noticed by PivIT Global specialists (a data center maintenance and repair company). They remind that excess heat cannot be used to heat houses and buildings if the data center is located in a remote and/or sparsely populated region. Moreover, such disposal requires the creation of an expensive infrastructure[69]. In addition to excess heat, greenhouse gases such as CO_2, CH_4 and N_2O (nitrous oxide) are released into the atmosphere. This is directly related to the operation and construction of data centers, where leaks of refrigerants used in cooling systems pose a particular threat to the environment. One must not forget about the use of diesel generators running if the main electricity is disconnected or during equipment testing, which occurs on a regular basis. There is also enormous consumption of water, given that most operating data centers use evaporative cooling, which releases heat into the environment (Thangam et al., 2024). For example, according to official data provided in a Google report, the total amount of greenhouse emissions associated with their activities increased by 48% in 2023 alone compared to 2019[70]. Independent investigations results look even less optimistic. The British Guardian presents its own analysis, according to which emissions from data centers owned by such technology giants as Google, Microsoft, Meta[71] and Apple from 2020 to 2022 could be 662% higher than officially registered ones; the underestimation is the result of imperfect accounting and certification systems[72].

The situation is aggravated by the fact that technologies are developing faster than their legal regulation, for example, in Norway[73]. In an attempt to bring the data centers under control, the Norwegian legislator supplemented Articles 3–7 of the Law “On Electronic Communications”. Now they stipulate mandatory registration of data center operators and a recommendation to use “the best available technical solutions, recognized standards,

---

68 Foslie, S. St. & Moen, O. M. (2021, March 16). This is how we reduce data centres carbon footprint. SINTEF. https://clck.ru/3NNkPD

69 4 Ways Data Center Heat Can Be Reused. (2024, March 26). Pivit Global. https://clck.ru/3NNkN7

70 Google Environmental Report 2024. https://clck.ru/3NNkM4

71 The organization is recognized as extremist, its functioning is prohibited in the territory of the Russian Federation.

72 Data center emissions probably 662 % higher than big tech claims. Can it keep up the ruse? (2024, September 15). https://clck.ru/3NNkL8

73 Andreassen, B. L. (2023, April 28). Scandinavian data centres: fewer jobs and less profit than forecast. Nordic Labour Journal. https://clck.ru/3NNkKJ

cost and usefulness of the measures applied"[74]. Assumingly, the following generally recognized standards and certificates applicable to DPCs are meant:

1. ISO 14001 is an internationally recognized standard for environmental management systems that contributes to the achievement of the UN SDGs on climate protection, responsible consumption and production, affordable and clean energy, etc.[75]. This international standard defines the requirements for an environmental management system that an organization can use to improve its environmental performance, namely, the organization's environmental policy development, environmental impact assessment, pollution control, potential risk assessment, and striving for continuous improvement of environmental performance[76].

2. ISO 50001 is an internationally recognized standard that implies the integration of energy management into the overall efforts of a certified company to improve environmental management[77].

3. LEED is an internationally recognized system for assessing the environmental friendliness of all types of buildings, which is based on solving the problem of climate change (protecting and restoring water resources, protecting biodiversity, reducing negative impacts on the planet's climate, etc.) and achieving the UN SDGs[78].

4. EU DC CoC is a European Data Centers Code of Conduct, which is a voluntary initiative, developed by the Joint Research Center and guiding data center owners and operators "in cost-effective reduction of energy consumption without compromising the critical function of their facilities"[79].

5. BREEAM is an internationally recognized certificate for assessing the environmental sustainability of buildings, where the ultimate goal is zero CO2 emissions by 2050[80].

6. Nordic Swan Ecolabel is an internationally recognized certificate based on the assessment of a product's full life cycle. Its purpose is to reduce the environmental impact during the production and consumption of goods[81]. Data centers do not belong to any of the Nordic Swan Ecolabel certified groups; however, the latter include materials used in construction and maintenance, as well as offices located on a data center territory[82].

7. EKOenergy label is an internationally recognized certificate of renewable electricity, heat, gas and cooling. It implies annual audit of the compliance of sold or used labeled

74 Ekomloven. (2024). Lovdata. https://clck.ru/3NNk43

75 ISO 14001:2015. https://clck.ru/3NNk5U

76 Environmental management systems – Requirements with guidance for use. https://clck.ru/3NNk6e

77 ISO 50001. https://clck.ru/3NNk7h

78 LEED rating system. https://clck.ru/3NNkB7

79 EU. European Code of Conduct for Energy Efficiency in Data Centres. https://clck.ru/3NNkDm

80 Achieve your net zero goals with BREEAM certification. https://clck.ru/3NNkEs

81 Why choose ecolabelling? Nordic Swan Ecolabel. https://clck.ru/3NNkGq

82 Criteria. Nordic Swan Ecolabel. https://clck.ru/3NNjgb

electricity with the EKOenergy criteria (location of solar panels and wind turbines; geothermal and offshore installations outside protected natural areas; production of hydroelectric power with the account the fish migration and the preservation of aquatic species' habitats, etc.) [83].

8. ASHRAE are international recommendations on heating, cooling and air conditioning[84].

9. Carbon Trust is an internationally recognized certificate aimed at reducing operational emissions and greenhouse gases. It requires companies to continuously finance relevant changes and disclose information about their implementation[85].

10. Provisions of European Union regulations, such as Commission Regulation 2019/424 of March 15, 2019, establish requirements for the environmental design of servers and data storage devices in accordance with Directive 2009/125/EC of the European Parliament and the EU Council[86].

At the same time, there is a trend in national legislations to mitigate the requirements. For example, the US plans to ease environmental restrictions for data centers in certain federal areas, where special power plants working on natural gas will be built. The latter will service massive data centers that consume at least 1 GW of electricity (the amount of energy consumed by a city with a population of 10,000,000 people) [87]. The corresponding order was signed by President Joe Biden in January 2025[88]. Russia is in the legal vanguard in this field. The Code of Rules 541.1325800.2024 "Buildings and structures of data processing centers. Design Rules" was adopted by the Russian Ministry of Construction to ensure compliance with the Federal Law No. 384-FZ dated December 30, 2009 "Technical Regulations on the Safety of Buildings and Structures". It complies with the requirements of Federal Laws No. 123-FZ dated July 22, 2008 "Technical Regulations on Fire Safety Requirements" and No. 261-FZ dated November 23, 2009 "On Energy Conservation and Energy Efficiency Improvement and on Amendments to Certain Legislative Acts of the Russian Federation". It also established mandatory general requirements for compliance with sanitary, epidemiological and environmental norms for the protection of human health, environment and adjacent buildings, energy saving and safety, and to basic engineering and technical systems of electrical and cold supply[89].

---

83 The EKOenergy ecolabel. EKOenergy. https://goo.su/GWlE2

84 Updated and Improved Standards Review Database. ASHRAE. https://clck.ru/3NNjkf

85 Net Zero transition planning and delivery. Carbon Trust. https://clck.ru/3NNjno

86 EU. (2019). Document 32019R0424. https://clck.ru/3NNjrF

87 Biden plan would encourage AI data centers on federal lands. (2024, December 19). The Washington Post. https://clck.ru/3NNjsv

88 Biden Wants Data Centers, Clean Energy on Federal Land by 2027. https://clck.ru/3NNjuN

89 Order of the Russian Ministry of Construction. Code of Rules 541.1325800.2024 of 23.12.2024. (2024). https://clck.ru/3NNje3

Environmental risks include land-use changes related to the construction, operation and provision of data centers with electricity. The construction of a data center and related infrastructure involves deforestation of large areas, as well as intake of water, often drinking water, which can lead to irreversible environmental consequences (Thangam et al., 2024). Green energy also involves active intervention in the natural landscape, from installing wind turbines and solar panels to drilling deep wells. This is done in Iceland, for example, where demand for electricity is higher than it can offer. The country developed the Icelandic Deep Drilling Project (IDDP), which implies increasing the number of deep wells in order to increase geothermal energy production[90].

Finally, projects for the construction of deep-sea marine cables also involve environmental risks. The entire life cycle of such a cable, including installation, maintenance and decommissioning, is associated with environmental impacts. These include: disturbance of the species' habitat, chemical and noise pollution, changes in electromagnetic fields, heat generation and other types of environmental damage (Taormina et al., 2018). It would be unfair to say that the laying and operation of deep-sea cables are not regulated in any way; nevertheless, both national and international laws pay much more attention to the cables protection than to environmental protection in the areas where they are laid. For example, Article 79 of the United Nations Convention in the Law of the Sea (1982) gives all countries the right to lay underwater cables and pipelines on the continental shelf, provided that "reasonable measures are taken to explore the continental shelf, exploit its natural resources, and prevent, reduce, and control pollution". It also obliges them to take into account previously laid cables and pipelines so as not to impair capabilities for their repair and maintenance[91]. The provisions of this Article serve as the basis of, for example, the internal US regulation. It this country, the National Association for Safety at Sea is responsible for issuing permits for laying underwater cables, including permits for placing underwater cables on the territories of national marine reserves[92].

The second group of risks associated with the growth of the data center industry in the Arctic states are social risks. Extraction of metals and minerals necessary for the data center operation, direct development and associated logging, construction of roads on pastures and lands of cultural value to the indigenous peoples living there, construction of hydro and other power plants – all this is a continuation of operational practices, including in relation to the Sami living in the Arctic regions. The legal protection of their pastures, as well as the realization of their rights as an indigenous people, is often only nominal and is reduced to zero with the growing demand for green electricity and the construction of data centers. At the same time, the Sami once again experience such

90 Moss, S. (2024, March 26). Iceland's AI moment. DCD. https://clck.ru/3NNjbz

91 United Nations Convention in the Law of the Sea. (1982). https://clck.ru/3NNjaX

92 Submarine Cables – Domestic Regulation. NOAA. https://clck.ru/3NNjXA

legacy of the past as land seizures and even forced resettlement. A bright example is the previously mentioned data center in Lulea, located on the pasture of Sami reindeer herders. Its construction was not even discussed with them[93]. The United Nations has already drawn attention to this problem, calling for the rights of indigenous peoples to be taken into account when developing deposits of "critically important minerals" associated with deforestation, water and soil pollution, loss of biodiversity and forced relocation of the indigenous population[94]. Promises to create a large number of jobs in places where indigenous peoples live and in remote areas of DPCs location are also not always true – instead of 30,000 jobs in Lulea, only 56 were created[95]. Other local residents also face deteriorating living conditions. Among the reasons are rising prices – for example, in Norway, which is attractive for its cheap and excess electricity, greenhouses in a number of regions were to be closed due to rising electricity prices[96]. All this leads to a rise of protests against the data centers construction around the world and, of course, in Northern Europe as their concentration point. Citizens believe that their rights and legitimate interests are violated, the quality of life is decreasing, and land is given to foreign technology companies as a priority[97].

Conclusions

The research results demonstrate that the construction of data centers and the development of the industry accompanying their operation and maintenance in the Arctic and Antarctic are associated with specific risks. The latter are determined by their geographical location, vulnerability of biological diversity and the ethnic composition of the population. These risks are divided into two main groups: environmental and social. Environmental risks are associated with the fact that the pace and scale of the occurring changes do not allow local ecological systems to adapt in a timely manner, and future changes cannot be adequately quantified (Robinson, 2022). At the same time, the slightest temperature fluctuations in their regions, as well as an increase in anthropogenic activity, can cause a chain reaction of irreversible climatic, sociological and economic changes throughout the planet[98]. Thus, the biological, food and physical

93 Sargysan, S. "Data Centers and Indigenous Sovereignty". https://clck.ru/3NNj6F

94 The UN urges to take into account the rights of indigenous peoples when developing deposits of "critically important minerals". (2025, April 23). UNO News. https://clck.ru/3NNj9a

95 Scandinavian data centres: fewer jobs and less profit than forecast. (2023, April 28). Nordic Labour Journal. https://clck.ru/3NNjBU

96 Andreassen, B. L. (2023, May 30). "Saving the environment" with liquid-cooled data centres. Nordic Labour Journal. https://clck.ru/3NNjEy

97 Tozzi, Ch. (2024, June 13). Why Communities Are Protesting Data Centers. Data Center Knowledge. https://clck.ru/3NNjLR

98 FAQ: Climate change in the Polar regions. SCRIPPS. https://clck.ru/3NNj2m

security of humanity is at risk. Social risks are associated with industrial development and land-use changes in the Arctic countries, namely with the ineffective implementation of the indigenous peoples' rights (the right to land, the right to health and a safe environment, the right to a decent life, the right to preserve and develop their own culture, etc.)[99]. In the scientific literature, this approach has already been called "green imperialism", i.e. the development of climate strategies in the interests of global elites with further marginalization of vulnerable communities. This implies the aggravation of centuries-old inequality and historical injustice. A striking example is the extraction of rare earth minerals and metals for the "green shift", disproportionately affecting the territories inhabited by indigenous peoples (Boretti, 2025). In the future, it may lead to new local and international conflicts. Today, we are already witnessing the territorial ambitions of the United States towards Greenland – a self-governing autonomy within the Danish Kingdom, the main population of which is Inuit.

The analysis of the international legal framework has revealed that, despite a large amount of declarations, conventions and agreements, it lacks specific acts to meet the current challenges of the new technological revolution. The latter requires increased mining, construction of modular and massive data centers, and laying of deep-sea cables. The Arctic territories are becoming increasingly attractive for that, while the Antarctic is turning into a center of scientific research on climate change. This, paradoxically, is associated with the growing harmful anthropogenic impact on the continent regions. Further industrial and anthropogenic development of the territories of the Arctic states and Antarctica will have global environmental and social consequences. At the same time, the national legislation shows low effectiveness in leveling the risks identified in this article. We can conclude that there is an urgent need to establish international regulation of the development of the data center industry in the Arctic and Antarctic. This system of regulation must include uniform requirements for certification and reporting, as well as responsibility for violations, taking into account the specifics of the environmental safety of these territories, their cultural and social characteristics.

References

Ali, S., Poto, M. P., & Murray, E. M. (2024). Arctic Vulnerability: Examining Biosecurity Risks Amidst Climate Change. In G. Panieri, M. P. Poto, E. M. Murray (Eds.), *Emotional and Ecological Literacy for a More Sustainable Society* (pp. 157–169). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-56772-8_8

Bannan, D., Ólafsdóttir, R., & Hennig, B. D. (2022). Local Perspectives on Climate Change, Its Impact and Adaptation: A Case Study from the Westfjords Region of Iceland. *Climate*, *10*(11), 169. https://doi.org/10.3390/cli10110169

Bargagli, R., & Rota, E. (2024). Environmental contamination and climate change in Antarctic ecosystems: an updated overview. *Environmental Science: Advances*, *3*(4), 543–560. https://doi.org/10.1039/d3va00113j

Boretti, A. (2025). Green imperialism a barrier to equitable progress in the hydrogen economy. *International Journal of Hydrogen Energy*, *105*, 137–147. https://doi.org/10.1016/j.ijhydene.2025.01.195

[99] UNO. (2007, September 13). UNO Declaration on the Rights of Indigenous Peoples. https://clck.ru/3NNiz2

Edwards, T. L., Nowicki, S., Marzeion, B. et al. (2021). Projected land ice contributions to twenty-first-century sea level rise. *Nature*, *593*(7857), 74–82. https://doi.org/10.1038/s41586-021-03302-y

Hughes, K. A., Convey, P., & Turner, J. (2021). Developing resilience to climate change impacts in Antarctica: An evaluation of Antarctic Treaty System protected area policy. *Environmental Science & Policy*, *124*, 12–22. https://doi.org/10.1016/j.envsci.2021.05.023

Markkula, I., Turunen, M., Rikkonen, T. et al. (2024). Climate change, cultural continuity and ecological grief: Insights from the Sámi Homeland. *Ambio*, *53*, 1203–1217. https://doi.org/10.1007/s13280-024-02012-9

Mathiot, P., & Jourdain, N. C. (2023). Southern Ocean warming and Antarctic ice shelf melting in conditions plausible by late 23rd century in a high-end scenario. *Ocean Science*, *19*(6), 1595–1615. https://doi.org/10.5194/os-19-1595-2023

Meredith, M. P., Inall, M. E., Brearley, J. A. et al. (2022). Internal tsunamigenesis and ocean mixing driven by glacier calving in Antarctica. *Science Advances*, *8*(47), eadd0720. https://doi.org/10.1126/sciadv.add0720

Orr, J.C., Kwiatkowski, L. & Pörtner, H. O. (2022). Arctic Ocean annual high in PCO2 could shift from winter to summer. *Nature*, *610*, 94–100. https://doi.org/10.1038/s41586-022-05205-y

Parmentier, F. J. W., Thornton, B. F., Silyakova, A., & Christensen, T. R. (2024). Vulnerability of Arctic-Boreal methane emissions to climate change. *Frontiers in Environmental Science*, *12*. https://doi.org/10.3389/fenvs.2024.1460155

Pecuchet, L., Mohamed, B., Hayward, A. et al. (2025). Arctic and Subarctic marine heatwaves and their ecological impacts. *Frontiers in Environmental Science*, *13*. https://doi.org/10.3389/fenvs.2025.1473890

Raimondi, L., Wefing, A.-M., & Casacuberta, N. (2024). Anthropogenic carbon in the Arctic Ocean: Perspectives from different transient tracers. *Journal of Geophysical Research: Oceans*, *129*(1), e2023JC019999. https://doi.org/10.1029/2023JC019999

Rantanen, M., Karpechko, A. Y., Lipponen, A. et al. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Commun Earth Environ*, *3*, 168. https://doi.org/10.1038/s43247-022-00498-3

Robinson, S. A. (2022). Climate change and extreme events are changing the biology of Polar Regions. *Global Change Biology*, *28*(20), 5861–5864. https://doi.org/10.1111/gcb.16309

Saunavaara, Ju., & Laine, A. (2021). Research, Development, and Education: Laying Foundations for Arctic and Northern Data Centers. *Arctic and North*, *42*, 145–169. https://doi.org/10.37482/issn2221-2698.2021.42.145

Siegert, M. J., Bentley, M. J., Atkinson, A., Bracegirdle, T. J., Convey, P., Davies, B., Downie, R., Hogg, A. E., Holmes, C., Hughes, K. A., Meredith, M. P., Ross, N., Rumble, J. & Wilkinson, J. (2023). Antarctic extreme events. *Frontiers in Environmental Science*, *11*, 1229283. https://doi.org/10.3389/fenvs.2023.1229283

Suopajärvi, L., Tikkanen, J., Edvardsdóttir, A. Engen, S., Inkilä, E., Iversen, A., … Ólafsdóttir, R. (2024). Geopolitical tensions framing different industries in the European Arctic: aquaculture, forestry, mining, and tourism in question. *Journal of Land Use Science*, *19*(1), 121–133. https://doi.org/10.1080/1747423X.2024.2357576

Taormina, B., Bald, J., Want, A., Thouzeau, G., Lejart, M., Desroy, N., & Carlier, A. (2018). A review of potential impacts of submarine power cables on the marine environment: Knowledge gaps, recommendations and future directions. *Renewable and Sustainable Energy Reviews*, *96*, 380–391. https://doi.org/10.1016/j.rser.2018.07.026

Terhaar, J., Tanhua, T., Stöven, T., Orr, J. C., & Bopp, L. (2020). Evaluation of data-based estimates of anthropogenic carbon in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *125*(6), e2020JC016124. https://doi.org/10.1029/2020JC016124

Thangam, D., Muniraju, H., Ramesh, R., Narasimhaiah, R., Muddasir Ahamed Khan, N., Booshan, S., Booshan, B., Manickam, T., & Sankar Ganesh, R. (2024). Impact of data centers on power consumption, climate change, and sustainability. In *Computational Intelligence for Green Cloud Computing and Digital Waste Management* (pp. 60–83). IGI Global. https://doi.org/10.4018/979-8-3693-1552-1.ch004

Turunen, M. T., Rikkonen, T., Nikula, A., Tuulentie, S., & Rautio, P. (2024). Between the local and the global? – reindeer herders' perspectives on land use challenges and conflicts in the Sámi homeland, Finland. *Journal of Land Use Science*, *19*(1), 134–149. https://doi.org/10.1080/1747423X.2024.2359606

Yuan, X., Liang, Y., Hu, X., Xu, Y., Chen, Y., & Kosonen, R. (2023). Waste heat recoveries in data centers: A review. *Renewable and Sustainable Energy Reviews*, *188*, 113777. https://doi.org/10.1016/j.rser.2023.113777

Živojinović, I., Elomina, J., Pülzl, H., Calanasan, K., Dabić, I., Ólafsdóttir, R., … Nygaard, V. (2024). Exploring land use conflicts arising from economic activities and their impacts on local communities in the European Arctic. *Journal of Land Use Science*, *19*(1), 186–210. https://doi.org/10.1080/1747423X.2024.2382676

Author information

Nataliya I. Shumakova – Senior Lecturer, Department of Constitutional and Administrative Law, South-Ural State University (National Research University)
Address: 76 Lenin prospekt, 454080, Chelyabinsk, Russia
E-mail: shumakovani@susu.ru
ORCID ID: https://orcid.org/0009-0004-6063-0650
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=59966654200
WoS Researcher ID: https://www.webofscience.com/wos/author/record/MTE-8168-2025
РИНЦ Author ID: https://www.elibrary.ru/author_items.asp?authorid=1211522

Conflict of interest

The author declares no conflict of interests.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – May 1, 2025
Date of approval – May 20, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:341.4:004.8
EDN: https://elibrary.ru/gcvuaw
DOI: https://doi.org/10.21202/jdtl.2025.15

Международные основы правового регулирования индустрии центров обработки данных в арктических государствах и Антарктике

Наталья Игоревна Шумакова

Южно-Уральский государственный университет (национальный исследовательский университет), Челябинск, Россия

Ключевые слова

Антарктика,
Арктика,
коренные народы,
международное право,
право коренных народов,
право,
центры обработки данных,
цифровые технологии,
экологическая безопасность,
экологическое право

Аннотация

Цель: критически оценить эффективность существующих международных правовых норм в условиях новых вызовов технологического прогресса, связанных с развитием индустрии центров обработки данных в арктических государствах и Антарктике.

Методы: методологическую основу исследования составляет комплекс специальных и общих методов научного познания, включая юридическую компаративистику, контент-анализ, дедукцию, индукцию, формально-логический метод и анализ документов. Автор уделяет внимание междисциплинарным подходам для объективной оценки экологических, социальных и правовых рисков, возникающих вследствие роста индустрии центров обработки данных в регионах с повышенной климатической и социальной уязвимостью.

Результаты: проведен анализ международных правовых актов, регулирующих деятельность центров обработки данных в полярных регионах. Выявлены ключевые риски, делящиеся на экологические (нестабильность локальных экосистем, неадаптивность к быстрым изменениям, риск потери биологического разнообразия и выбросы парниковых газов) и социальные (маргинализация и нарушение прав коренных народов, утрата традиционных культур и образа жизни, рост социальной напряженности). Указана неизбежность появления новых конфликтов и вызовов вследствие недостаточной эффективности национальных и международных механизмов регулирования. Констатирована необходимость создания специализированных международных правовых инструментов, учитывающих специфику экологической безопасности полярных территорий.

Научная новизна: статья впервые дает комплексную картину совокупных рисков и недостатков действующего международного регулирования индустрии центров обработки данных в арктических государствах и Антарктике. Проведен детальный сравнительный анализ нормативной базы, показана несоответственность применения принципов «мягкого права» в полярных регионах в эпоху четвертой технологической революции. Обосновано требование о создании новых сертификационных и отчетных процедур на всем жизненном цикле центров обработки данных с учетом правового и культурного контекста.

Практическая значимость: результаты работы ориентированы на совершенствование международной и национальной политики в сфере регулирования индустрии центров обработки данных, разработку стандартов сертификации и отчетности, эффективных в условиях климатических, социальных и экономических особенностей арктических стран и Антарктики. Направлены на минимизацию негативного влияния антропогенных факторов и обеспечение баланса между индустриальным развитием и сохранением уникальных природных и культурных ландшафтов.

Для цитирования

Шумакова, Н. И. (2025). Международные основы правового регулирования индустрии центров обработки данных в арктических государствах и Антарктике. *Journal of Digital Technologies and Law*, *3*(3), 369–396. https://doi.org/10.21202/jdtl.2025.15

Список литературы

Ali, S., Poto, M. P., & Murray, E. M. (2024). Arctic Vulnerability: Examining Biosecurity Risks Amidst Climate Change. In G. Panieri, M. P. Poto, & E. M. Murray (Eds.), *Emotional and Ecological Literacy for a More Sustainable Society* (pp. 157–169). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-56772-8_8

Bannan, D., Ólafsdóttir, R., & Hennig, B. D. (2022). Local Perspectives on Climate Change, Its Impact and Adaptation: A Case Study from the Westfjords Region of Iceland. *Climate*, *10*(11), 169. https://doi.org/10.3390/cli10110169

Bargagli, R., & Rota, E. (2024). Environmental contamination and climate change in Antarctic ecosystems: an updated overview. *Environmental Science: Advances*, *3*(4), 543–560. https://doi.org/10.1039/d3va00113j

Boretti, A. (2025). Green imperialism a barrier to equitable progress in the hydrogen economy. *International Journal of Hydrogen Energy*, *105*, 137–147. https://doi.org/10.1016/j.ijhydene.2025.01.195

Edwards, T. L., Nowicki, S., Marzeion, B. et al. (2021). Projected land ice contributions to twenty-first-century sea level rise. *Nature*, *593*(7857), 74–82. https://doi.org/10.1038/s41586-021-03302-y

Hughes, K. A., Convey, P., & Turner, J. (2021). Developing resilience to climate change impacts in Antarctica: An evaluation of Antarctic Treaty System protected area policy. *Environmental Science & Policy*, *124*, 12–22. https://doi.org/10.1016/j.envsci.2021.05.023

Markkula, I., Turunen, M., Rikkonen, T. et al. (2024). Climate change, cultural continuity and ecological grief: Insights from the Sámi Homeland. *Ambio*, *53*, 1203–1217. https://doi.org/10.1007/s13280-024-02012-9

Mathiot, P., & Jourdain, N. C. (2023). Southern Ocean warming and Antarctic ice shelf melting in conditions plausible by late 23rd century in a high-end scenario. *Ocean Science*, *19*(6), 1595–1615. https://doi.org/10.5194/os-19-1595-2023

Meredith, M. P., Inall, M. E., Brearley, J. A. et al. (2022). Internal tsunamigenesis and ocean mixing driven by glacier calving in Antarctica. *Science Advances*, *8*(47), eadd0720. https://doi.org/10.1126/sciadv.add0720

Orr, J.C., Kwiatkowski, L. & Pörtner, H. O. (2022). Arctic Ocean annual high in PCO2 could shift from winter to summer. *Nature*, *610*, 94–100. https://doi.org/10.1038/s41586-022-05205-y

Parmentier, F. J. W., Thornton, B. F., Silyakova, A., & Christensen, T. R. (2024). Vulnerability of Arctic-Boreal methane emissions to climate change. *Frontiers in Environmental Science*, *12*. https://doi.org/10.3389/fenvs.2024.1460155

Pecuchet, L., Mohamed, B., Hayward, A. et al. (2025). Arctic and Subarctic marine heatwaves and their ecological impacts. *Frontiers in Environmental Science*, *13*. https://doi.org/10.3389/fenvs.2025.1473890

Raimondi, L., Wefing, A.-M., & Casacuberta, N. (2024). Anthropogenic carbon in the Arctic Ocean: Perspectives from different transient tracers. *Journal of Geophysical Research: Oceans*, *129*(1), e2023JC019999. https://doi.org/10.1029/2023JC019999

Rantanen, M., Karpechko, A. Y., Lipponen, A. et al. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Commun Earth Environ*, *3*, 168. https://doi.org/10.1038/s43247-022-00498-3

Robinson, S. A. (2022). Climate change and extreme events are changing the biology of Polar Regions. *Global Change Biology*, *28*(20), 5861–5864. https://doi.org/10.1111/gcb.16309

Saunavaara, Ju., & Laine, A. (2021). Research, Development, and Education: Laying Foundations for Arctic and Northern Data Centers. *Arctic and North*, *42*, 145–169. https://doi.org/10.37482/issn2221-2698.2021.42.145

Siegert, M. J., Bentley, M. J., Atkinson, A., Bracegirdle, T. J., Convey, P., Davies, B., Downie, R., Hogg, A. E., Holmes, C., Hughes, K. A., Meredith, M. P., Ross, N., Rumble, J. & Wilkinson, J. (2023). Antarctic extreme events. *Frontiers in Environmental Science*, *11*, 1229283. https://doi.org/10.3389/fenvs.2023.1229283

Suopajärvi, L., Tikkanen, J., Edvardsdóttir, A. Engen, S., Inkilä, E., Iversen, A., … Ólafsdóttir, R. (2024). Geopolitical tensions framing different industries in the European Arctic: aquaculture, forestry, mining, and tourism in question. *Journal of Land Use Science*, *19*(1), 121–133. https://doi.org/10.1080/1747423X.2024.2357576

Taormina, B., Bald, J., Want, A., Thouzeau, G., Lejart, M., Desroy, N., & Carlier, A. (2018). A review of potential impacts of submarine power cables on the marine environment: Knowledge gaps, recommendations and future directions. *Renewable and Sustainable Energy Reviews*, *96*, 380–391. https://doi.org/10.1016/j.rser.2018.07.026

Terhaar, J., Tanhua, T., Stöven, T., Orr, J. C., & Bopp, L. (2020). Evaluation of data-based estimates of anthropogenic carbon in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *125*(6), e2020JC016124. https://doi.org/10.1029/2020JC016124

Thangam, D., Muniraju, H., Ramesh, R., Narasimhaiah, R., Muddasir Ahamed Khan, N., Booshan, S., Booshan, B., Manickam, T., & Sankar Ganesh, R. (2024). Impact of data centers on power consumption, climate change, and sustainability. In *Computational Intelligence for Green Cloud Computing and Digital Waste Management* (pp. 60–83). IGI Global. https://doi.org/10.4018/979-8-3693-1552-1.ch004

Turunen, M. T., Rikkonen, T., Nikula, A., Tuulentie, S., & Rautio, P. (2024). Between the local and the global? – reindeer herders' perspectives on land use challenges and conflicts in the Sámi homeland, Finland. *Journal of Land Use Science*, *19*(1), 134–149. https://doi.org/10.1080/1747423X.2024.2359606

Yuan, X., Liang, Y., Hu, X., Xu, Y., Chen, Y., & Kosonen, R. (2023). Waste heat recoveries in data centers: A review. *Renewable and Sustainable Energy Reviews*, *188*, 113777. https://doi.org/10.1016/j.rser.2023.113777

Živojinović, I., Elomina, J., Pülzl, H., Calanasan, K., Dabić, I., Ólafsdóttir, R., … Nygaard, V. (2024). Exploring land use conflicts arising from economic activities and their impacts on local communities in the European Arctic. *Journal of Land Use Science*, *19*(1), 186–210. https://doi.org/10.1080/1747423X.2024.2382676

Сведения об авторе

Шумакова Наталья Игоревна – старший преподаватель кафедры конституционного и административного права, Южно-Уральский государственный университет (национальный исследовательский университет)
Адрес: 454080, Россия, г. Челябинск, пр. Ленина, 76
E-mail: shumakovani@susu.ru
ORCID ID: https://orcid.org/0009-0004-6063-0650
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=59966654200
WoS Researcher ID: https://www.webofscience.com/wos/author/record/MTE-8168-2025
РИНЦ Author ID: https://www.elibrary.ru/author_items.asp?authorid=1211522

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.87 / Международное право
Специальность ВАК: 5.1.5 / Международно-правовые науки

История статьи

Дата поступления – 1 мая 2025 г.
Дата одобрения после рецензирования – 20 мая 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.

Research article
UDC 34:004:343.721:004.8
EDN: https://elibrary.ru/smgmxq
DOI: https://doi.org/10.21202/jdtl.2025.16

Neurohacking in the Digital and Artificial Intelligence Age: Legal Aspects of Protecting Neural Information

Diogo Pereira Coelho
University of Seville, Seville, Spain

Keywords

artificial intelligence,
brain-computer interface,
cybercrime,
digital technologies,
law,
metaverse,
neurocrime,
neurohacking,
neurosecurity,
neurotechnologies

Abstract

Objective: to contribute to the concept of neurocrime; to study the current and future risks from the viewpoint of cybersecurity in the context of digitalization and artificial intelligence development.

Methods: the study uses a critical and descriptive analysis of the relationship between cybercrime and neurocrime. It provides a conceptual distinction between the brain-computer interface and its use and describes the differences between neural and mental manipulation. The legal autonomy of crimes against mental integrity in relation to crimes against physical integrity is investigated. The methodological framework includes the analysis of existing prototypes of neurocrimes based on a four-phase brain-computer interface cycle and the study of the features of neurohacking in the context of the metaverse and artificial intelligence technologies.

Results: the study revealed the essential characteristics of neurohacking as the misuse of neural devices to gain unauthorized access to and manipulate neural information. Four main types of brain-computer interface applications subject to neurohacking are identified: neuromedical applications, user authentication systems, video games, and smartphone-based applications. The modalities of neurohacking were established at each phase of the brain-computer interface cycle: manipulations at the stage of neural information input, measuring and recording of brain activity, decoding and classifying neural information, as well as at the stage of the result output. The specific threats of neurohacking in the era of digitalization are analyzed, including immersive attacks and human joystick attacks in the metaverse.

Scientific novelty: for the first time, a comprehensive differentiation of the concepts of neurocrime and cybercrime was carried out, highlighting their specific legal consequences. The author proposed a classification of neurocrimes based on the four-phase cycle of the brain-computer interface. The study substantiated the need to distinguish mental integrity as an independent object of legal protection, different from the protection of physical integrity. For the first time, the features of neurohacking in the context of the metaverse and artificial intelligence technologies were investigated, including the analysis of new types of attacks and threats to neurosecurity.

Practical significance: the study results are important for the development of legal regulation in the field of cybersecurity and the corresponding regulations. The identified types of neurocrimes and their classification can help to create a specialized legislation on the protection of neural data and mental integrity. Practical recommendations on ensuring the neurosecurity of brain-computer interfaces are in demand in medical practice, video game industry, authentication systems, and for the development of smartphone applications.

For citation

Coelho, D. P. (2025). Neurohacking in the Digital and Artificial Intelligence Age: Legal Aspects of Protecting Neural Information. *Journal of Digital Technologies and Law, 3*(3), 397–430. https://doi.org/10.21202/jdtl.2025.16

Contents

Introduction

There are those who believe that illicit access to or manipulation of neural information is not and never will be feasible in the way that is often assumed and feared[1]. The main reason for this is the limited understanding of the neural code, i.e. the language through which the brain encodes and processes information. For this to be possible, it would be necessary to decode the neural code in order to achieve a certain result (access or manipulation) and, among the billions of neurons that exist in the human brain, to determine which specific one to stimulate[2]. Although it is currently possible to make predictions about which region of the brain to stimulate, identifying the exact neuron still appears to be a major challenge. In addition, the neuron responsible for a particular function in the brain of a given subject may not be the same in the brain of another subject[3]. On the other hand, some argue that it is not possible to influence a subject's behavior by stimulating just a single neuron, because brain function depends on the coordinated activity of complex neuronal circuits, involving sets of hundreds or billions of neurons. The coordinated stimulation of large networks of specific neurons,

1 Fields, R. D. (2022). Hacking the brain: More fantasy than reality. The UNESCO Courier. Should we be afraid of neuroscience? (p. 9). UNESCO. https://clck.ru/3Nkhbr

2 Ibid.

3 Ibid.

with a view to imposing targeted and specific behavior for the purposes of manipulation and mental control, appears to be practically impossible[4].

It turns out that, in general, everything is considered a system, including the human brain. And like all systems, the human brain also seems to be the target of "hacking". Human beings themselves are considered[5] natural life hackers[6]. Card counting in blackjack[7] is considered a hack. Most sports are usually targets for hacking purposes. In F1[8], teams try to find new ways to modify the design of vehicles that are not expressly forbidden by the regulations. Gerrymandering[9] is considered a hack used in politics. Also in the financial and business world, there are several hacking methods used. From entrepreneurs to financial institutions, most of the players try to find loopholes in the system (i.e. the law), essentially in order to gain an advantage over competitors. Specifically, they exploit situations that are not expressly prohibited, but represent intentional (or not) subversions of the system. Both Uber[10], and Airbnb[11], among other large technology companies, initially violated the rules imposed by the legal systems of multiple jurisdictions. The concept of "obstruction" thus appears to be a hack as old as the concept of "gap in the law" or "gap in the system" itself. It is probably as old as human civilization, since people themselves seem to be able to be hacked. In this context, the human brain presents itself as a system. Specifically, a system whose optimization has resulted from continuous interaction with the environment over millions of years, having been developed to survive and, above all, to reproduce.

---

4 Ibid.

5 Schneier, B. (2021). The Coming AI Hackers. Belfer Center for Science and International Affairs. Harvard Kennedy School. https://clck.ru/3Nkhe4

6 The English expression "hacker" is usually translated to other languages as "computer pirate". However, using this kind of translated word appears to be inaccurate. A "hacker" can act like a "pirate" and a "pirate" can act like a "hacker", but the two definitions do not seem to depend on each other, nor do they seem to be equivalent. Rather, they seem to be purely complementary. A hacker exploits systems, whether on a computer, on a phone, through personal interaction or simply in any other aspect of human life. And this, as a rule, in a lawful way. Hackers (or "crackers") illicitly try to decode/ crack a system, either for fun or to obtain a certain result or advantage. Regardless of semantic issues and terminological confusion, this text will mostly use the term "hacker", essentially because it is the most well-known expression and also to make it easier to read. See Ribeiro, J. B. (2019, 12 Fevereiro). 'Hacker' vs 'Pirata Informático': a riqueza de uma definição perdida na tradução. SH/FTER. https://clck.ru/3NkhkJ

7 Keating, S. (2022). How a magician-mathematician revealed a casino loophole. BBC. https://clck.ru/3Nkhnx

8 Straw, E. (2022, February 22). F1's new philosophy in combatting design loopholes. The Race. https://clck.ru/3Nkhq9

9 Ax, J. (2023). North Carolina court allows partisan gerrymandering. Reuters. https://clck.ru/3Nkhrs

10 Henley, J. (2017, September 29). Uber clashes with regulators in cities around the world. The Guardian. https://clck.ru/3Nkht2

11 Neubauer, I. L. (2019, August 30). Countries that are cracking down on Airbnb. The New Daily. https://goo.su/UcOHh

Cognitive hacking thus appears to be a powerful tool[12] in relations between individuals, with the manipulation technique known as "social engineering" standing out. Within cybercrime itself, the only novelty lies in the use of technology for hacking purposes, because just like the human brain, computers are also systems. Over the last few decades, we have seen hacking methods adapt to the computerization of traditional systems. This computerization seems to have changed hacking methods in three different ways: scale, scope and speed. Firstly, it has amplified and extended the nature of hacks, thereby increasing their scale and scope. Next, the growing number of software and hardware developments has allowed systems to evolve faster than initially anticipated. Computer speed kept pace with this development, which resulted in an increase in the speed of hacking methods[13]. With the evolution from Web 1.0 to Web 2.0 and Web 3.0 and, more recently, Web 4.0, new disruptive technologies have emerged[14] and the use of digital or artificial computer systems is expanding at an increasingly rapid pace (Ienca & Haselager, 2016)[15]. Ultimately, with the evolution of cybercrime, this use will focus on the human brain and mind itself, specifically in the form of "neurocrime" and "neurohacking" (also commonly referred to as "brain hacking"). In this scenario,

12 Schneier, B. (2021). The Coming AI Hackers. Belfer Center for Science and International Affairs. Harvard Kennedy School. https://clck.ru/3Nkhe4. By way of example, many of the powerful social systems that form the basis of society, such as democracy and the market economy, among others, depend on the decisions that people make. This process can be the target of cognitive hacking in multiple ways. Starting with social communication. Personalized according to our preferences and behaviors, modern advertising represents a kind of mass hacking of the human brain, specifically of the conscious psychic process, including the previous unconscious state. Not to mention disinformation (often disseminated by the press itself), which represents a hack of the common understanding of reality. The repeated use of terms such as "terrorism" or "cyberterrorism" in the media and in politicians' speeches also represents a hack of the cognitive system, essentially with a view to convincing people that this is a greater threat than it really is and thereby causing fear and misrepresenting risk assessment.

13 Schneier, B. (2021). The Coming AI Hackers. Belfer Center for Science and International Affairs. Harvard Kennedy School. https://clck.ru/3Nkhe4

14 The term "disruptive" is relatively recent and, in other words, means causing a transformation in the standards, models or technologies already established in the market. In other words, it describes a technological innovation, product or service with so-called "disruptive" characteristics, because it is diverse, revolutionary, innovative or never thought of or applied in that specific context. See Dufloth, R. (2017). Novas tecnologias e o futuro do profissional do Direito. Mgalhas. https://clck.ru/3Nm3jy

15 In general, the technologies associated with web 1.0, web 2.0, web 3.0 and web 4.0 allow users to interact directly with data and systems and can be used for a wide variety of daily or even professional tasks. For example, while GPS systems help with geolocation and spatial navigation, portable devices monitor bodily processes such as heart rate, calorie intake and weight loss. Still by way of example, personal computers help with cognitive tasks such as arithmetic calculations, writing and memory, blockchain-based digital assets allow international transactions of value in a matter of minutes or even seconds, and generative artificial intelligence systems to produce text, images or videos instantly and innovatively. See Nath, K. (2022). Evolution of the Internet Web 1.0 to Metaverse: The Good, the Bad and the Ugly. Research Gate. https://clck.ru/3NkiAo

the increase in scale, scope and speed of hacking methods will be increasingly noticeable.

The aim of this text is to contribute to the study and initial framing of a subject whose understanding will never be sufficient, not least because of the high level of legal assets at stake.

1. From cybercrime to neurocrime

1.1. Connection between cybercrime and neurocrime

There is no agreed definition of cybercrime. The terms "cybercrime", "computer crime", "computer-related crime" or "high-tech crime" are used frequently, but randomly. Whether at international, European or even national level, there is no consensus on the expression, definition, typology or classification of cybercrime (Rodrigues, 2009; Vasconcelos Casimiro, 2000). There is no concept of "cybercrime" or "computer crime" expressly established in Portuguese legislation[16]. There is also no uniformly settled concept in literature and jurisprudence (Venâncio, 2011). The lack of uniformity lies in the fact that the term "cybercrime" covers, in a generic and abstract way, a range of crimes committed using information and communication technologies. This term includes both classic criminal actions and new types of crime.

According to the European Commission, cybercrimes are "criminal acts committed using electronic communications networks and information systems or against such networks and systems"[17], and can be divided into three forms. Firstly, traditional forms of criminal activity, but using the Internet (and identity theft or phishing methods) to commit crimes (such as computer fraud or spoofing). These traditional forms also include the international electronic trade in drugs, weapons and endangered animal species. Secondly, the online publication of illegal content, such as material inciting terrorism, violence, racism and xenophobia or the sexual abuse of minors. Finally, crimes exclusively committed on electronic networks, which represent new and often "large-scale" crimes that were "unknown in the pre-internet age". In the latter case, criminal agents attack systems or entire information infrastructures and even confidential State information (which constitutes a national threat). Still according to the European Commission, these attacks can be carried out through "botnets" (network of robots), i.e. criminal agents distribute

[16] As an example, see the case of Portugal. In addition to the types of crimes provided for in Law no. 109/2009, of September 15, which approves the Portuguese Cybercrime Law. (https://clck.ru/3Nkxa7), there are also other types of crimes of this nature provided for in the Portuguese Penal Code and in various other separate legal sources.

[17] European Commission. (2007). Towards a general policy on the fight against cyber crime. Communication from the Commission to the European Parliament, the Council and the Committee of the Regions. https://clck.ru/3NkiNh

"malware" which, in turn, when transferred, transforms the user's computer into a "bot". Once the network or information infrastructure is infected, it is used to commit crimes without the knowledge of the respective users. In short, computer crime constitutes any act in which the computer or similar technology serves as a means to achieve a criminal objective, and the computer or similar technology may represent a merely symbolic target of that act or even represent the object of the crime (Marques & Martins, 2006). It is therefore necessary to distinguish between "computer crime", in which information technology is the target of the crime, and "crime committed using computer means", in which information technology is the means of executing the crime.

In this logic, whenever crime is committed using neural interfaces and, in addition to representing a physical threat to users, it can also have a profound impact on their behavior and self-perception, everything indicates that we are facing a "neurocrime"[18]. The problem associated with the misuse of information technologies in the context of neurotechnology is particularly critical, as this type of technology applies (directly or indirectly) to the brain, one of the most important organs in the human body. The human brain not only contributes significantly to life processes, such as reproduction and maintenance of life, but also provides consciousness, perception, the ability to think, discernment, memory and language. Furthermore, it still has enormous importance in human behavior and self-perception as a sensitive being or individual endowed with emotions and sensitivity (Ienca & Haselager, 2016). Neurocriminality thus seems to constitute any act in which the human brain and/ or mind serve as a means to achieve a criminal goal, and the human brain and/ or mind can represent a merely symbolic target of this act or represent the object of the crime. Therefore, a distinction shall be made between "neurocriminality", in which the brain and/or mind directly or indirectly constitute the target of the crime, from "crime committed using neural and/or purely mental manipulation", in which brain and/ or mind constitutes the means of executing the crime. As such, as we will see in point 3 of this chapter, a distinction shall also be made between neural manipulation and purely mental manipulation. In addition, neurocrime may not involve direct access to the brain or stored information, but only indirect access. For example, the functions of the neural device may simply be limited, modified or deregulated. With the current advances in neural engineering technologies (mainly marketed in the health sector), this scenario appears to be increasingly a reality. In this scenario, the perpetrator of the neurocrime can affect the victim's brain indirectly, since the neural system is not directly accessed or significantly manipulated during the attack. Nonetheless, it may affect the victim's mental state in a significant way, as the neurocrime may have limited or constrained their conduct,

[18] A term commonly used in the case of criminal activities that exploit neural devices.

generated an emotional response in the form of panic, fear, or psychological disturbances, and/or left traumatic memories. It is worth mentioning that, in certain circumstances, the perpetrator and the target of the crime may be confused. For example, users with mental instability may choose to damage their neural devices in order to attempt suicide (Ienca & Haselager, 2016). That said, in a broad sense, the concept of neurocrime seems to be able to be defined as the crime of offense against the mind of a person or a group of people, committed using neural and/or purely mental manipulation via a neural device, and with the intention of, directly or indirectly, causing physical or mental harm, including reputational and/or property damage.

In the field of neurocrime and neural stimulation, there are currently two types of neural devices considered particularly critical. On the one hand, there are brain stimulators, especially deep brain stimulation (DBS) and transcranial direct current stimulation (tDCS) systems. On the other hand, brain-computer interfaces (BCI) stand out. Both types of device allow direct access to neural computing, although in different ways (brain simulation vs. reading brain activity). In addition, both types of devices are available not only as medical technologies, but also as products marketed to users considered healthy. As such, both types of devices raise multiple concerns in terms of "neurosecurity". The fact is that, to date, BCIs have been the main neural devices used for hacking purposes (even with experimental evidence and in the context of real situations), which is why they are the only ones to be explored in this article (Ienca & Haselager, 2016).

1.2. Concept of brain-computer interface and use cases

Unlike mere neurostimulators (electronic devices similar to cardiac pacemakers), BCIs are not used to stimulate the brain, but rather to establish a direct communication link which, by bypassing the peripheral nervous system and muscles, allows users to control an external computer exclusively through brain activity (Ienca & Haselager, 2016; Vallabhaneni et al., 2005). BCIs were first developed in the field of clinical medicine and as a therapeutic or medical assistance technology for neurological patients. In a clinical context, BCI applications are used to repair, assist, or augment motor, cognitive, or sensory functions in patients suffering from neurological disorders that precisely affect motor development and/ or cognitive and sensory functions, including spinal cord injuries, strokes, and neurological motor diseases such as amyotrophic lateral sclerosis (ALS) and muscular dystrophy (Ienca & Haselager, 2016).

BCI can be distinguished into two types: invasive and non-invasive. While invasive BCI records brain activity through the surgical implantation of electrode arrays in the central nervous system or through a mere direct connection, non-invasive BCI records brain activity through electrodes placed on the outside of the skull, i.e. through neuroimaging

technologies such as electroencephalography (EEG) and electromyography (EMG). In both cases, a direct interaction is established between the user's brain and the neural device. As a rule, this interaction consists of a cycle of four phases (Ienca & Haselager, 2016; Van Gerven et al., 2009; Bernal et al., 2022)[19]. The first phase consists of the input of neural information (i.e. the creation of specific brain activity) by the user in response to a given stimulus (whenever the BCI user wants to perform a certain mental task or achieve a certain cognitive state). The second phase consists of measuring and recording brain activity. In this phase, the user's brain activity patterns are detected, measured, and recorded by the interface during the cognitive process or the execution of a certain mental task. In the third phase, the raw neural data (neural information) resulting from the second phase is decoded in order to assess its main characteristics and classify it. After decoding and classification, the data is translated into the production of a certain result (output), i.e. the production of the result expected by the user. In general, in this fourth phase, the output consists of carrying out the action initially intended, desired, or considered beneficial to the user through the control of applications connected to the BCI. Controllable applications include powered devices such as electric wheelchairs or even robotic prosthetic limbs, sensory devices and other types of software and hardware applications (including mobile applications and cell phones). Once each of these phases has been completed, in principle, the user can see the result of the previous phase and thus start the next phase (Ienca & Haselager, 2016). An example of this is Gert Jan Oskam, who, after a motorcycle accident and being a paraplegic for over a decade, is now able to walk again using BCIs (albeit imperfectly for the time being)[20].

Today, BCI applications are available not only in clinical settings, but also for the general public (Mochan et al., 2025). Multiple commercial applications of EEG-based BCI devices have entered the market and are increasingly popular for both video games and everyday activities (Ienca & Haselager, 2016)[21]. In the electronic telecommunications industry, there are also BCIs available on the market. Firstly, certain mobile applications, such as "Xwave" (launched over 10 years ago!), allow certain types of earphones to establish a direct connection with compatible iPhone cell phones and, through this connection, record brain

[19] In this regard, it should be noted that BCIs can be classified based on its levels of invasiveness. While non-invasive systems require electrodes to be applied to the scalp, invasive systems require a surgical procedure to place electrodes inside the skull, either on the surface of the brain or even inside the brain.

[20] Whang, O. (2023). Brain Implants Allow Paralyzed Man to Walk Using His Thoughts. https://clck.ru/3Nkxhk

[21] As an example, see the websites of the companies Emotiv Inc. and Neurosky Inc., both pioneers in the commercialization of non-invasive, intuitive and accessible BCIs for video games, interactive televisions or non-manual control systems, respectively accessible at: https://clck.ru/3NkiYN & https://clck.ru/3NkiZy. See also Gordon, L. (2020, December 16). Brain-controlled gaming exists, though ethical questions loom over the tech. The Washington Post. https://clck.ru/3Nkicf

waves and frequencies (Ienca & Haselager, 2016)[22]. The same is currently happening in the arms industry, as several BCI applications are being developed (Ienca & Haselager, 2016; Czech, 2021). For example, the US Defense Advanced Research Projects Agency (DARPA) funds a wide range of BCI projects, essentially aimed at restoring behavioral or neural functions in soldiers and also improving the training and performance of those same soldiers and even secret service agents (Ienca & Haselager, 2016; Kotchetkov et al., 2010; Miranda et al., 2015). Based on the vast potential of brain control via neural computing devices in terms of utility and basic functionality, it is predicted that BCI will gradually replace the keyboard, touch screen, computer mouse and even voice assistance technology, as consumers may prefer to interact directly with computers (Ienca & Haselager, 2016; Yuan et al., 2010; Radu, 2024). However, as we will see in the next points and chapters, while the potential benefits of the foreseeable mass commercialization of certain clinical and non-clinical BCI technology applications appear to be significant and well-studied, the very serious (and perhaps irreparable) risks associated with neurosecurity remain largely unexplored. In order to understand these risks, a distinction must first be made between neuronal and purely mental manipulations.

1.3. Distinction between neural and purely mental manipulations

After more than 2,000 years of philosophical argument and the most recent impressive scientific advances in artificial intelligence (AI) and data science, the problem of mind-brain dualism persists and continues to defy a unanimous solution (Bublitz & Merkel, 2014; Moulin, 2022). According to Maria Fernanda Palma, the absolute denial of mind-brain dualism, or the reduction of the phenomena of consciousness to brain states, can have a significant impact on the real behavioral basis on which legal responsibility is founded, particularly in relation to voluntary action or true capacity of individuals to be considered guilty under Criminal Law[23]. In the event that conscience does not temporarily precede the decision, the free and conscious decision as a criterion of responsibility or greater responsibility in the figure of intention is necessarily questioned. In this sense, the absolute self-determination of the human being as a source of responsibility is questioned, when it leads to the assumption that the decision intelligent can be calculated like an algorithm and imitated by its mathematical processes. While neuroscience seems to propose a naturalistic reduction of the mind and states of consciousness, AI and data science suggest that the functioning of the brain can be conceived as an autonomous biological process, in which mental states and their relationship to human behavior can be

22 PLX Devices Inc. XWave – Mind Interface Introduction and Teaser. 2010. https://clck.ru/3NkifC

23 Fernanda Palma, M. (2021–2022). Ciberneurodireito. Powerpoint Presentation. Short Course in Criminal Law on Artificial Intelligence and Artificial Intelligence in Criminal Law. FDUL. IDPCC/ CIDPCC, 4.

reproduced. This perspective unexpectedly reopens the dualist proposal, treating the brain as just one support among other possible bases for human behavior. As such, and based on the associated sciences, the Law aims to examine whether a specific mental state has a causal or correlated cerebral existence, and whether it necessarily implies a certain behavior, as well as to what extent it is controllable[24].

According to Jan Christoph Bublitz & Reinhard Merkel, all mental phenomena are, in one way or another, connected to brain activity. In this sense, it is argued that the legislator should be cautious when describing the mind-brain relationship as a dualistic relationship, not least because it is difficult to identify changes in the mental state without also identifying certain changes at the cerebral (neural) level. It is also argued that mental states are not only correlated with a certain brain state, but are also caused or made conscious on the basis of a certain physical state. However, whenever a more precise and concrete description of this correlation is considered, multiple problems arise, as there is no consensus on the correlation with the causes or the process of "awareness" (Bublitz & Merkel, 2014). As such, despite the current tendency in psychiatry to classify any mental disorder as a brain disorder in the strict sense, there are those who argue that mental injuries may not necessarily be confusable with physical bodily injuries. In principle, it is not the brain that "decides", "suffers" or feels the "moral damage". On the contrary, everything indicates that we are dealing with mental states/ processes of people and not exactly with qualities of physical objects. For normative purposes related to the concepts of "harm" or "disorder/ dysfunction", everything indicates that the mind and brain deserve individual attention (Bublitz & Merkel, 2014). By way of example, mental disorders/ dysfunctions seem to result from certain psychological functions or in relation to certain social norms, and not exactly from electrochemical brain processes. In this case, it seems that we are dealing with mental and behavioral phenomena that cannot be described purely in terms of neuroscience. By way of example, it seems that "depression" is a specific mental symptom and suffering from depression depends exclusively on the display of that symptom. Even if it were known (which it is not) that all symptoms of depression are strongly correlated with chemical imbalances at the neurotransmitter level, the distinction between mental and brain dysfunction would persist (Bublitz & Merkel, 2014). In this logic, it has been argued that mental damage should not be treated in the same way as brain damage. Otherwise, the advances and developments of modern criminal law will be disregarded and the Roman concept of iniuria, which was seen as a summary notion of any and all offenses committed against the person, will be considered. The literature has thus been advocating the definition of mental states worthy of legal protection and the introduction of specific normative provisions that penalize interference with mental integrity, rather than the protection of physical integrity being adapted and expanded.

[24] Fernanda Palma, M. (2021–2022). Ciberneurodireito. Powerpoint Presentation. Short Course in Criminal Law on Artificial Intelligence and Artificial Intelligence in Criminal Law. FDUL. IDPCC/ CIDPCC, 4, Pp. 6–14.

While brain integrity should cover physical interventions, i.e. brain damage (regardless of the mental consequences), mental integrity should cover mental interventions, i.e. mental damage and regardless of the brain consequences (Bublitz & Merkel, 2014).

In this regard, it is also argued that mental integrity should also encompass protection against mental manipulations/interventions, such as the provocation of emotions, the manipulation of preferences and decision processes, the optimization of non-consensual neurological development, the manipulation of memory, the manipulation of will or willpower, among other cognitive and emotional phenomena (Bublitz & Merkel, 2014). In all these cases, the manipulations/ interventions restrict mental capacities or alter preferences and will formation. However, to date, none of these manipulations/ interventions seem to be, in general, adequately covered by the rules that currently protect physical integrity and mental health. Not least because all of these manipulations/ interventions only cause psychological changes, and do not seem to meet the requirements to be considered physical harm. Since the victim is not affected by any brain damage or experience of physical pain or discomfort, the unlawfulness of these manipulations/ interventions seems to stem from their purely mental effects (Bublitz & Merkel, 2014). With the entry into the digital and artificial age, everything indicates that in the future, in the event of litigation related to brain and mental integrity, neuroscientific evidence and proof will be increasingly important in ascertaining and discovering the truth (Shen, 2013). For this reason, and regardless of whether the denial of the mind-brain dualism persists, before we move on to the analysis of neurohacking per se, it is still necessary to make some considerations regarding the current deliberation of the legal autonomy of crimes against the mind in relation to crimes against physical integrity.

1.4. Considering the legal autonomy of crimes against mental integrity in relation to crimes against physical integrity

As explained in the previous point, neuroscience, AI and data science not only question assumptions about the position of the mind in the natural world, but also instigate a rethinking of its role in the normative world. It turns out that the law currently offers unilateral protection, as it systematically protects the body and the brain and only fragmentarily the mind and mental states. As explained in the previous point, the fundamental question is to what extent it is possible to legitimately intervene in the minds and mental states of other subjects. With the commercialization and mass implementation of neural technologies with the ability to intervene in the mind and detect mental activity, it is argued that the law should introduce autonomous and individual legal protection of mental integrity (Bublitz & Merkel, 2014; Abegão Alves, 2020). This scientific-philosophical and legal debate basically focuses on two aspects: (i) considering the empirical and conceptual autonomy of the mental in relation to the physical; (ii) considering the legal autonomy of crimes against mental integrity in relation to crimes against physical integrity. According to the Research Project “Crimes

Against the Mind", developed by the Institute of Criminal Law and Criminal Sciences of the Faculty of Law of the University of Lisbon (after the publication of the pioneering study by Bublitz & Merkel, 2014), which aims to fill this gap in legal thinking, the search for the limits of the legitimate alteration of the mental states of others has not been carried out by jurists and legal thinkers, so that the reality to be regulated remains one step ahead of the Law. Also according to this project, the most recent scientific discoveries demand the attention of the Law, not only from the point of view of the aggressor/ criminal, but also from the point of view of the victim, and the legal debate should arise both in relation to the problems of free will and the foundation of criminal responsibility that these discoveries raise, and also in relation to the legal assets to be protected. It is this second facet of the relevance of neuroscience to law that has yet to be explored. In this sense, the current aim is to deepen the scientific-philosophical/ legal debate on this subject, both at national and international level. Specifically, the aim is to problematize the key issues, proposing solutions and possible paths for future investigations that will inevitably involve the intersection of different fields of knowledge.

2. From neurocrime to neurohacking

2.1. Neurocrime prototypes usually referred to as neurohacking

Over the last few years, multiple examples of prototype neurocrimes usually referred to as neurohacking have been identified. These include the hijacking of wireless limb prostheses, the malicious reprogramming of neural stimulation therapy (i.e. unauthorized wireless changes to the device's configuration in order to generate certain brain stimuli) and the unauthorized interception of brain implant signals in order to obtain private neural information (Ienca & Haselager, 2016)[25]. In principle, all these examples can only be carried out using neural devices that allow a direct connection to the brain to be established, such as tDCS and especially the increasingly developed BCI (Denning et al., 2009). These are the prototypes of neurocrime that are usually referred to as neurohacking and, as we will see in the next few points, the way they are carried out is very similar to computer hacking in the context of cybercrime as explained in the first point of the first chapter on the connection between cybercrime and neurocrime (Ienca & Haselager, 2016).

2.2. Concept of neurohacking

In a broad sense, neurohacking seems to consist of the abusive and malicious use of neural devices in order to illicitly obtain and possibly manipulate neural information (Ienca & Haselager, 2016). Strictly speaking, neurohacking seems to consist of a neuro-attack carried out through neural devices, through which the perpetrators gain illicit access

[25] This last example describes a specific neurocriminal phenomenon in which the attack is not simply aimed at disrupting the neural device, but at obtaining unauthorized access to private information.

to neural information which, in turn, can be manipulated in order to control the cognitive process or the execution of a certain mental task of the user of the device. Once the neural device has been accessed, it is used to commit crimes with or without the user's knowledge.

2.3. Neurohacking based on types of BCI applications

2.3.1. Scope

As explained so far, BCIs can be intercepted to detect hidden autobiographical information from users and with a significantly high accuracy rate (Ienca & Haselager, 2016; Rosenfeld et al., 2006; Rosenfeld, 2011). It has been shown that once intercepted, BCIs can reveal private and confidential information about users, such as their PIN codes, bank and credit card details, date of birth, home address and even the faces of people they know (Ienca & Haselager, 2016). The science fiction future described in the present article therefore seems to be anything but fiction. In fact, such a future, in which subjects are able to access and manipulate the neural information of other subjects, is not approaching at a rapid pace, as it is a current reality (Mochan et al., 2025). In this scenario, and as we will see in the next sections of this chapter, as well as in Chapter III, unless the design and functional characteristics of current neural devices, which are still under development, have strong neurosecurity measures, their misuse and malicious use could imply serious and grave risks in terms of public safety (Ienca & Haselager, 2016)[26].

2.3.2. BCI applications that could be the target of neurohacking

According to the first point of this Chapter II, neurocrime prototypes can only be carried out through neural devices that make it possible to establish a direct connection with the brain, and BCI applications are the main targets for hacking. It is possible to distinguish between four main types of BCI applications that could be targeted for neurohacking: (i) neuromedical applications; (ii) user authentication; (iii) video games and entertainment; and (iv) smartphone-based applications. For each of these applications, possible attack scenarios are currently being studied, as well as the respective neurosecurity measures to be applied (Mochan et al., 2025). In fact, for some of the neurohacking activities based on types of BCI applications, there is already evidence and experimental proof and even in a real context (Li et al., 2015). As we will see in the next point, within the scope of these four main types of BCI applications, neurohacking can, in principle, take place in any of the different phases of the BCI cycle described in point two of Chapter I.

[26] Martinovic, I. et al. (2012). On the feasibility of side-channel attacks with brain-computer interfaces. USENIX security symposium. https://clck.ru/3NkoD5

2.4. Neurohacking modalities based on the four-stage cycle of the BCI

2.4.1. Scope

As mentioned in the previous point, and taking into account the types of BCI applications that already exist (Mochan et al., 2025) and also, theoretically, those that may exist in the (very) near future, the next sub-points will cover, also theoretically, the various types of neurohacking based on the four-phase BCI cycle. Specifically, the input phase of neural information, the phase of measuring and recording brain activity, the phase of decoding and classifying neural information and, finally, the output phase (Ienca & Haselager, 2016).

2.4.2. Manipulation of the neural information input phase

In this form of neurohacking, the hacker attacks the BCI user at the moment of inputting neural information, i.e. in the first phase of the cycle. The input of neural information can be manipulated by changing the stimuli communicated to the BCI user. For example, neurohackers can pre-select target stimuli in order to trigger a specific response in the user and thus facilitate access to their neural information. This type of neural malware has some similarities to spyware on a computer, as it aims to collect information about the user, send it to another entity, and/ or ensure control over a computer or other IT device without the user's permission or consent (Ienca & Haselager, 2016). In this case, the malware used for neurohacking has the particularity of being able to extract information directly from the signals emitted by the brain, and is therefore commonly known as "brain-spyware". In the future, there seem to be several possible mobile and portable applications of brain-spyware for the purposes of neurocrime and, it seems, activities such as password cracking, identity theft, phishing and other types of neural scams will become increasingly common (Ienca & Haselager, 2016). For the time being, deciphering the signals emitted by the brain with a level of precision and speed that is comparable to computer hacking does not seem to be possible outside of an experimental environment (and, in this case, to a limited extent). Given the limitations in terms of deciphering signals emitted by the brain, as well as the current degree of maturity of the market, the reward does not seem worth the risk for today's hacker. However, with technological advances and the rapid expansion of the BCI applications market, it seems that neural information will be increasingly valued and also decipherable and readable (Ienca & Haselager, 2016).

2.4.3. Manipulation of the brain activity measurement and recording phase

In the phase of measuring and recording brain activity, the hacker attacks the BCI user with the aim of, without the latter's consent or will, producing a result (output) different from that expected due to the regular functioning of the neural device. The form of the attack may differ depending on its specific purpose, with three main purposes in mind: cracking raw neural data from BCIs (i.e. neural information), disrupting the functionality of the BCI, and hijacking the BCI. In general, cracking raw data can result in various criminal

activities aimed at limiting, harming, or taking advantage of certain neural device user behavior. Attacks aimed at disrupting the functionality of the BCI can occur whenever the hacker intends to manipulate the measurement of brain activity in order to confuse, overdrive, or delay the functions of the BCI application. Hijacking can occur whenever the hacker aims to monitor and alter the BCI communication channel in order to decrease or even replace the user's level of control over the BCI application. During hijacking, the neural device receives orders that diverge from its user's intentions or desires and benefit the hacker. By way of example, this type of neurohacking could involve disrupting a particular speech production device based on BCI in order to silence the user, or even hijacking a wheelchair in order to define the user's route (Ienca & Haselager, 2016).

2.4.4. Manipulation of the decoding and classification phase of neural information

Neurohacking of the decoding and classification phase involves manipulating the production of the result (output) intended and expected by the user as a function of the normal processing of the BCI device. This can be done in three different ways: (i) by introducing noise in order to make the decoding process unnecessarily challenging; (ii) by interfering with the neural learning and memory mechanisms (machine learning) in order to manipulate the classification of brain waves; or (iii) by replacing the waves sent by the BCI to the output device. Each of these methods has advantages and disadvantages. For example, adding noise appears to be the easiest method to perform and the most difficult to detect compared to other methods. However, with this method, the hacker's chances of achieving the desired result are reduced. On the other hand, although the other two methods are the most difficult to execute and the easiest to detect, at the same time, they also offer the most control over the BCI system. As with the manipulation of the brain activity measurement and recording phase, the hacker appears to be able to manipulate the decoding and classification phase in order to hijack the BCI device. In this type of attack, the aim appears to be not only to limit or take control of the device away from the user, but also to replace it. By successfully hijacking the system, the hacker appears to be able to gain partial or total control over the BCI device, while limiting or eliminating the user's control. In this scenario, the hacker appears to be able to monitor, alter or insert messages into the BCI communication channel (Ienca & Haselager, 2016).

2.4.5. Manipulation of the output production phase

This modality occurs whenever the hack aims to alter the output perceived by the user at the end of each BCI cycle. In this modality, the neurohacker aims to manipulate the user's perception of immediately previous actions or their own self-perception resulting from cognitive states generated by the BCI. The criminal motive behind this type of hack would be, without the user's permission, to induce specific cognitive states or actions in the user's subsequent cycle (or in each subsequent cycle) for the hacker's benefit. By way of example,

the neurohacker appears to be able to carry out a kind of "neurophishing". In this sub-modality, the user may be induced by the hacker to enter a certain password or other type of authentication information before the originally intended process can begin or continue. In this scenario, it seems possible to subject the user to certain traumatic experiences. The criminal activities that fall under this category, among others, include scamming, phishing, identity theft and harm to physical or mental integrity (Ienca & Haselager, 2016).

3. Neurohacking in the digital and artificial age

3.1. Concept of digital and artificial age

It is commonly accepted that the information age consists of a historical period (between the 50s and 70s of the 20th century) that saw the transformation of traditional industries, established during the industrial revolution, to an economy centered on information (and communication) technology. Today, with the beginning of the 21th century and the economy moving towards digital and artificial technologies, it seems that we have entered a new historical period, specifically the digital and artificial age. As mentioned in the introduction, with the evolution of Web 1.0 to Web 2.0 and Web 3.0 and, more recently, to Web 4.0, new disruptive technologies have emerged and the use of digital and artificial computer systems is expanding at an increasingly rapid pace. Among these new disruptive technologies, recent advances in the sector of immersive technologies, the metaverse, and virtual worlds, or even in the sector of AI and data science, stand out. In the first sector, essentially in terms of virtual (VR), augmented (AR), mixed (MR), and extended (RX) reality, and also in terms of spatial computing and digital sensory interaction. In the second sector, mainly in terms of generative AI systems and robotics[27]. Both these sectors promise to revolutionize the way society interacts with technology (and vice versa). Not only do these types of technologies aim to improve the user experience in terms of efficiency, but they also have the features and characteristics to drive multiple sectors forward at breakneck speed (Ford, 2016)[28] and at a rate that exceeds all predictions[29], including in sectors associated with other disruptive technologies such as cloud and edge computing, big data, quantum computing, brain-computer interfaces, distributed ledger technologies

[27] See GPT-4. https://clck.ru/3NkoQK; Bing AI. https://clck.ru/3NkoTY; Gemini. https://clck.ru/3NkoV7. See also the advances in terms of humanoid robots, specifically: Boston Dynamics. https://clck.ru/3NkoY3; Tesla. https://clck.ru/3Nkxk8

[28] In this regard, it should be noted that the best-known way of measuring the progress of computer processing power is "Moore's Law". Gordon E. Moore predicted that every 18 months, the number of transistors on chips would increase by 100%. However, information technology is exceeding what Moore's Law itself predicted. Unlike hardware, for example, which has seen significant increases in computer memory capacity and the amount of digital information that can be transmitted over fiber optic cables, software is seeing the effectiveness of certain algorithms grow at a rate that exceeds all predictions (Ford, 2016).

[29] Coelho, D. P. (2023). Os recentes avanços no setor da IA são uma benção ou uma maldição? Observador. https://clck.ru/3NkpED

(such as blockchain), the internet of things (IoT), smart cities, facial recognition, robotics, among others.

Imagine these technologies reaching the potential widely predicted (Dwivedi et al., 2022)[30]. In the next 15 to 20 years, everything will be directly connected. New worlds and new countries will be (re)created in digital form in the metaverse[31]. Smart cities will be connected via IoT, with CCTV's[32] in every corner and highly equipped with facial recognition and intelligent sensory systems[33]. AI-based drones and humanoid police robots will patrol the streets and buildings[34]. Universities will be mostly AI-based[35]. Most organizations will not even have workers or physical spaces, being autonomous, digital and based on AI[36]. Businesses will be digital, in the metaverse and developed by autonomous digital and artificial companies[37]. AI-based industrial robots will produce all kinds of goods[38]. Workplaces will be virtually all digital[39]. People will spend practically their entire day (and night) in the increasingly developed metaverse[40]. People will be able to choose to step out into the physical environment in the form of holograms, making it possible to attend meetings and workspaces and, even to "move" freely anywhere on the planet[41]. All this, of course, without leaving home. The line separating virtual reality from augmented reality and mixed reality will become smaller. The same goes for the line separating physical

30 See also Coelho, D. P. (2023). Ano 2050: Era Digital. Observador. https://clck.ru/3NkpRV; Chayka, K. (2021). We already live in Facebook's metaverse. The New Yorker. https://clck.ru/3NkjZb

31 See Woodward, W. (2024). Backup nations: countries making digital twins to mitigate natural disasters. Nesta. https://goo.su/uVSUy; Widlund J. (2023) Singapore's First Country-Scale Digital Twin and The Future of Digital Open Data. https://clck.ru/3NkxqG

32 Also known as "closed circuit television" or "surveillance cameras".

33 Davis, D. (2021). Facial Recognition And Beyond: Journalist Ventures Inside China's 'Surveillance State'. NPR. https://clck.ru/3NkjET

34 See Jarecki, J., Wilson, N., & Trevellyan, K. (2024). Vermont police are using drones more than ever. Here's what that means. Vermont Public. https://clck.ru/3NkjGh ; Chen, H. (2023). Robôs com mais de 2 metros de altura integram força policial de Singapura em aeroporto. CNN Brazil. https://goo.su/W2x0

35 Carroll, M. (2024). UK's first 'teacherless' AI classroom set to open in London. Sky News. https://clck.ru/3NkjLK

36 Smith, T. (2024). Profitable, AI-powered companies with no employees to arrive 'next year'. Sifted. https://clck.ru/3NkjNX

37 Eckert, T., & Cigaina, M. (2023). The metaverse: A new space for business. SAP. https://clck.ru/3NkjPh

38 See Ping, Ch. (2019) Robots to wipe out 20 million jobs around the world by 2030: Study. https://clck.ru/3Nkxtp; Semuels, A. (2020). Millions of Americans Have Lost Jobs in the Pandemic – And Robots and AI Are Replacing Them Faster Than Ever. Time. https://clck.ru/3Nkxxo

39 Hoover, A. (2024) The Metaverse Was Supposed to Be Your New Office. You're Still on Zoom. https://clck.ru/3Nky4S

40 Steele, C. (2022). People Are Spending More Time Online-and They're Not Happy About It. PC Mag. https://clck.ru/3NkzDa

41 See Atkinson, E., & Meyer, M. (2022). Meeting in the Metaverse: The Future of Work?. University of Denver, Podcast. News. https://clck.ru/3Nkpxk. Verdict. In the metaverse, holograms offer more options than avatars. (2022, June 15). https://clck.ru/3Nkq2b

reality from digital reality. Central bank digital currency will be the only legal tender[42]. In everyday tasks, people will be assisted by portable and wearable technologies (in many cases invisible)[43] or even by humanoid domestic robots based on AI[44]. In the decisions they make, they will be assisted by AI through voice assistants or even through neural interface technology, with a view to merging human consciousness with AI in a kind of symbiosis between human and machine[45]. In general, each time people establish a connection with the metaverse, a neural connection will be initiated[46]. Digital sensory interaction will make people feel increasingly comfortable connected to the metaverse, and this is the reality that the new generations will know best. Because of this digital life, people will always be surrounded by cameras, microphones and interface systems. Even from birth, embryonic development will take place in artificial incubators[47]. In fact, many people will choose to create and cultivate emotional or loving relationships with AI-based humanoid domestic robots[48]. Domestic animals themselves will be replaced by artificial pets[49]. Everything will be digital or artificial.

As such, thoughts themselves will not be safe, because whenever a connection is established with the metaverse and/ or any type of neural interface technology is activated, the human being will be an open book. In this scenario, the use of computer systems is not limited to the social, professional and economic spheres, but extends to the psychological and biological spheres. Inevitably, the dizzying speed of development in all these sectors, as well as their combination, could result in an equally dizzying increase in the scale, scope and speed of cybercrime, neurocrime and, consequently, neurohacking methods. As such, and as we will see in the scope of this chapter III, the study and investment in cybersecurity and neurosecurity could become increasingly relevant (Pooyandeh et al., 2022).

---

42 Michel, N. (2024, June 17). CBDCs Are Instruments Of Control-And They're Here. Forbes. https://clck.ru/3Nkq5h

43 From Wearables to Implantables: The Rise of Invisible Technologies. https://clck.ru/3NkyPr

44 See Reuters. (2024). A humanoid robot to help you around the house. https://clck.ru/3NkqCB; Schwartz, R. (2024). Is the world ready for Tesla's new domestic robots? The Week. https://clck.ru/3NkqE8

45 See Brodsky, S. (2024, August 27). AI voice assistants evolve, promising deeper interaction. IBM. https://clck.ru/3NkqHZ; Niemeyer, K. (2024, August 3). Elon Musk says Neuralink could help humans compete with AI: 'Let's give people superpowers'. Business Insider. https://clck.ru/3NkqLa

46 How BCI can elevate the AR/VR experience. https://clck.ru/3NkyXe

47 Zimmer, K. (2021, March 30). The Ultimate Incubator: The Brave New World of Bionic Babies – Artificial placentas could improve the survival odds of premature infants. IEEE Spectrum. https://goo.su/PE2QYzM

48 See Travers, M. (2024, March 24). A Psychologist Explains Why It's Possible To Fall In Love With AI. Forbes. https://clck.ru/3NkqXn; Chow, A. (2023). AI-Human Romances Are Flourishing-And This Is Just the Beginning. Time. https://clck.ru/3NkyiH

49 World Economic Forum. (2023). Moflin, an AI pet, responds like a real animal. https://clck.ru/3Nkqgf

3.2. Concept of metaverse

In 1992, science fiction author Neal Stephenson introduced the term “metaverse” in his 1992 cyberpunk novel “Snow Crash”. In this work, a 3D virtual world is presented in which people, represented as avatars, could interact with each other and with artificially intelligent agents. In 2003, this initial concept of a metaverse (still a long way from the concept that is currently being idealized) was first implemented in the game Second Life and even had some success[50]. The term “metaverse” is formed by combining the Greek prefix “meta”, which can be translated as “beyond” or “transcendence”, and the suffix “verse”, which comes from the word “universe”. We are thus dealing with a world beyond the universe (Bernal et al., 2022). As such, this concept aims to represent a virtual (digital) world which, despite coexisting with physical reality (through augmented reality), allows us to overcome the physical limitations of the real world, such as space and time. In this digital environment (which, it seems, will be increasingly valued), multiple users aim to interact exactly as they do in real life, using an avatar that represents their digital alter ego or even their digital identity (Bernal et al., 2022). In a broad sense, the concept of the metaverse therefore consists of a space or set of virtual and shared spaces (commonly called digital or virtual worlds or environments) where users, represented by digital avatars, can access and interact in a multidimensional way via their headsets (among other possible accessories). In other words, instead of simply viewing the content, users can immerse themselves in digital content through their digital representations[51].

The main and basic technologies that currently make up the metaverse and virtual worlds are immersive technologies (such as virtual, augmented, mixed and extended reality, BCIs, and sensory interaction systems), 3D modeling and reconstruction technologies, spatial and edge computing, AI and data science, IoT, and distributed recording technologies (Pooyandeh et al., 2022) [52]. Unlike today’s virtual and/ or augmented reality technologies, which are mostly used for electronic games or to replace the keyboard, touch screen or even the computer mouse, the technology we want to achieve could be used to simulate practically any situation associated with the physical world. From carrying out professional

[50] Second Life is a video game launched in 2003 that allows users to enjoy a “second life” in the virtual world. Users can assume any identity and play any role. Specifically, they can take on the role of an avatar in a virtual world that can be explored to meet other users, to take part in individual and/ or group activities, and so on, just as they would in real life. See the Second Life video game website. https://clck.ru/3Nm4mQ

[51] Pereira Coelho, D. (2021). Metaverse: should regulators be more attentive than ever? Observador. https://clck.ru/3Nm549

[52] See also Tucci, L. (2024, March 22). What is the metaverse? An explanation and in-depth guide. TechTarget. https://clck.ru/3Nkqqk

activities, to attending virtual concerts[53] or even just enjoying some time with friends, there will be multiple possibilities for interaction. The ultimate objective is, therefore, to eliminate the boundaries between the physical world and virtual reality, allowing users to interact with virtual objects through the physical world and vice versa, thus having the possibility of processing any information or value in real time.

Using distributed ledger technology, users can buy and sell non-fungible crypto-assets through fungible crypto-assets within the metaverse. In fact, within the scope of a "blockchain-based virtual world" operating on the basis of a "virtual economy", crypto-assets issued using blockchain technology, in addition to allowing the digital representation of fungible financial products, also allow the digital representation of non-fungible non-financial products, whether they are hard assets, i.e. tangible and physical, or soft assets, i.e. intangible or digital goods. In these terms, the possibilities are virtually limitless, with some arguing that the metaverse appears to be the next generation of the internet (Pooyandeh et al., 2022). One way or another, the metaverse seems to be at least an evolution of the internet, with a dominant focus on social interaction. As the metaverse develops and the number of users increases, it seems that more and more personal information will be at risk, including neural information, as we'll see in the next few points (Pooyandeh et al., 2022).

3.3. Digital sensory interaction in the metaverse

Nowadays, in addition to interacting with the smartphone or tablet screen (among others) through touch, sensory interaction with digital environments is usually limited to hearing and sight, i. e. in total to three of the five senses of the human body traditionally known. In the near future, it seems that interaction may include more basic senses, with their perception increasingly similar to that of the physical world. In 2013, Google published the search engine "Google Nose" and offered a service that allowed users to find the product they were looking for through the sense of smell[54]. Although this service was prepared as a kind of April Fool's joke and for the specific conditions of that day, it was considered a success and, at the time, it was clear that users seemed to be ready to take interaction with the internet to the next level[55]. Since then, user-friendly digital sensory interaction systems have been developed to include the sense of smell in the usual interaction with digital environments. By way of example, users of digital environments are intended to have the means to enjoy experiences where they can smell a perfume before purchasing it via

[53] Simões Ferreira, R. (2022, Desember 29). With holograms or in the metaverse, how digital has already reinvented 'live'. Jornal de Notícias. https://clck.ru/3Nkqum

[54] See the Google Nose Beta website. https://clck.ru/3Nkqw4

[55] Nordyke, K. (2023). Google's April Fools' Joke: Search and Smell (Video). The Hollywood Reporter. https://goo.su/UBRW9

e-commerce. Still as an example, the aim is for users to have the means to breathe in the smell of the sea while leaving home as if they were on the beach or even feel its humidity on their skin. The same goes for taste and the sensation of flavor. In 2020, at Meiji University in Japan, a prototype was developed (dubbed the "Tasting Device") that allows the user to experience different flavors from a device adapted to touch with the tongue[56]. In this sense, the concepts of "augmented human being" or "augmented human intelligence" will be increasingly used and common in the context of users' sensory interaction with digital environments.

Within the scope of the metaverse, mobile and wearable devices such as, for example, virtual and/ or augmented reality headsets, in addition to including sensors for the user to detect movement or sound, may also include other types of sensors. Specifically, virtual reality systems are made up of inertial measurement units and include accelerometers, gyroscopes and magnetometers. Time, breathing and light sensors are also included. Augmented reality systems, on the other hand, can detect the user's location and what they see or hear, and most headsets are equipped with time-of-flight (ToF) sensors, vertical cavity surface-emitting lasers (VCSELs), binocular depth sensors, and optical sensors for structural monitoring. Both systems can also include audio-related sensors such as directional microphones, as well as thermal sensors, touch sensors, and front and rear video cameras. The touch sensor can be used to exchange information between humans and machines in the form of a human-machine interface, with the tactile stimulus being activated by this type of sensor (as, for example, in a touchpad). Most of these sensors are used in the context of IoT in an industrial or clinical context, in drones, humanoid robots, among others (Pooyandeh et al., 2022).

However, the equipment associated with the metaverse, which is currently commercialized en masse, still has multiple limitations (both in terms of hardware and software). The overwhelming majority do not seem to be developed enough to offer a considerably immersive metaverse. The perception of sensations in the physical world still seems to be better than in the digital world. Consequently, the overwhelming majority of metaverse platforms do not manage large volumes of user data, which in turn means that we are not close to mass adoption either. The apparent failure of the "Apple Vision Pro" headset, which aims to replace the keyboard, touchscreen and computer mouse, seems to be the best example of this[57]. In this context, and as we will see in the next section, the literature identifies BCIs as the key technology for achieving complete integration

[56] Grad, P. (2020). Digital device serves up a taste of virtual food. TechXplore. https://clck.ru/3Nkr7r

[57] Mitchell, A. (2024, November 12). Apple's Vision Pro flop: Company scales back production of $3,500 VR headset amid lackluster sales, customer complaints. New York Post. https://clck.ru/3Nm5Ka

between the user and the metaverse in the medium to long term (Bernal et al., 2022)[58]. In addition, the development of cutting-edge sensors and hardware, as well as other types of equipment associated with the metaverse, also seems to contribute to the mass adoption of this new model of sensory interaction with the internet[59].

3.4. Brain-computer interface in the metaverse

Neuralink Corp.[60], a nanotechnology company partly owned by Elon Musk[61], has developed a type of BCI that requires neurosurgery to implant an integrated circuit (chip) in the user's brain, which is a concept that both intrigues and discourages many potential consumers. This type of BCI makes it possible to establish bidirectional interaction with the brain, which includes both the neural mechanisms of learning and memory, as well as neurostimulation. In general, although they can be used to restore the ability to speak, write and even walk, they are still viewed with some suspicion[62]. Nevertheless, the literature has studied and contributed to the discussion of the "state of the art", in particular by comparing current BCI to certain virtual and/ or augmented reality devices that constitute a kind of EEG cap. Because it causes ergonomic problems, it was found that the rapid development of virtual and/or augmented reality headsets (among other products) that incorporate too many EEG sensors to monitor brain regions is impeded. It has been argued in the literature that virtual and/ or augmented reality headsets that mitigate the effects of noise resulting from the processing of brain waves seem to be the most suitable for incorporating EEG sensors (Orlosky et al., 2021). It has also been argued that a combination of resources and technologies, including virtual and/ or augmented reality, digital avatars, sensory interaction systems and BCIs, seems to be able to make the metaverse increasingly pervasive and immersive in everyday life. If this is the case, it could reshape the social experience of space and time. The combination of the metaverse

---

58 Consequently, whether we are aware of it or not, we are currently witnessing a rapid and drastic transformation of business and commercial etiquette, both in terms of products and production methods, as well as in types of services and the way they are executed. By way of example, the subconscious or dreams of users themselves do not seem to be safe from marketing campaigns or even propaganda. There are sensory interaction systems that seek to affect the content of the user's dreams through brain stimuli carried out before or during sleep. Specifically, they try to induce the user to view a particular product or service during their dreams.

59 Genser, J., Damianos, S., & Yuste, R. (2024). Safeguarding Brain Data: Assessing the Privacy Practices of Consumer Neurotechnology Companies. The Neurorights Foundation. https://clck.ru/3NkrK7

60 The BCI device developed by Neuralink consists of a small probe containing more than 3,000 electrodes attached by flexible wires that are thinner than a human hair. This device can monitor the activity of 1,000 brain neurons. A "neurosurgical robot" was also built and it can insert 192 electrodes into the brain every minute. See Galeon, D. (2017, November 22). Experts: Artificial Intelligence Could Hijack Brain-Computer Interfaces. Can we prevent AI from hacking into the human brain? Futurism. https://clck.ru/3NkrMh

61 See the Neuralink website. https://clck.ru/3NkrP3

62 Hall, S. B., & Baier-Lentz, M. (2022, February 7). 3 technologies that will shape the future of the metaverse – and the human experience. The World Economic Forum. https://clck.ru/3NkrRf

and BCIs presents arguments for generating new forms of social interaction and interoperability, making communication between the physical world and the digital world ever faster, more effective and efficient, but also more transparent (Dwivedi et al., 2022). In addition to its usefulness in the medical context, the combination of the metaverse and BCIs is also useful in other types of contexts. Specifically, it allows users to control certain objects (tangible or intangible, such as certain robotics products or digital avatars) with their minds, mental spelling, authentication with brain waves or simply enjoying video games or other entertainment (Bernal et al., 2022). This combination can also be used for cognitive assessment, emotional control and increased cognitive performance. Current literature also explores the feasibility of using BCIs to allow direct communication between the brains of different subjects, using both neural mechanisms for learning and memory, as well as neurostimulation (Bernal et al., 2022). However, despite the notorious evolution of BCIs over the last few decades, their full implementation in metaverse scenarios has not yet been studied in the depth it deserves. There still seem to be some open challenges. Firstly, it seems necessary to broadly analyze how BCIs can contribute to the metaverse. Secondly, it seems necessary to measure the performance of these systems and identify the trends and challenges that BCI presents when applied in a metaverse scenario. Last but not least, it also seems necessary to identify the problems, limitations and risks associated with the use of BCIs in the metaverse (Bernal et al., 2022).

3.5. Neurohacking in the metaverse

There are many similarities between the internet and the metaverse when it comes to cybersecurity challenges such as hacking into accounts, phishing, malware, etc. Despite the differences in terms of infrastructure, the metaverse (web 3.0 & web 4.0) presents new types of cybercrime that differ from those that occur on traditional websites (web 2.0). As the use of crypto-assets and central bank digital currencies expands, hackers will be increasingly interested in cracking the metaverse (Pooyandeh et al., 2022). In this sense, monitoring the metaverse and detecting attacks on new platforms will be more complicated than on traditional platforms. In line with what has been exposed so far in the previous points, with the commercialization and mass adoption of products related to the metaverse and virtual worlds, the scope of hackers' activities will increase substantially. Among the main associated risks are the "immersive attack", i.e. a new type of attack in a virtual environment that focuses on the malicious manipulation of a given device in order to physically or mentally harm or disturb the user. Also noteworthy is the "human joystick" attack. This attack consists of controlling users immersed in virtual and/or augmented reality systems within the metaverse, without their knowledge or authorization, in order to move their physical body to another location within the physical world. With the combination of BCIs, especially those used for neurostimulation,

the attacks aim to over-stimulate the target brain regions or inhibit them, thereby interrupting regular brain activity. The damage caused by this type of threat appears to be able to even recreate the effects of neurodegenerative diseases, although more studies in this regard still need to be carried out (Bernal et al., 2022)[63].

3.6. Brain-computer interface based on artificial intelligence

AI has contributed to advances in the analysis and decoding of neural activity, and has even boosted the BCI sector. Over the last decade, a wide range of AI-assisted or even purely AI-based BCI applications have emerged. These "smart" BCIs, including motor and sensory BCIs, have shown remarkable clinical success. In addition to improving the quality of life of paralyzed patients, they have expanded the athletic ability of ordinary people, and accelerated the evolution of robots and neurophysiological discoveries. However, despite technological advances, there are still several challenges in relation to long periods of training and learning (machine learning), producing results (outputs) in real time and also measuring and recording brain activity in the scope of operation of this new type of BCIs. As explained in the previous point (and, in general, within the scope of this chapter III), it seems that there is still a need for more studies in this direction (Zhang et al., 2020).

3.7. Neurohacking and artificial intelligence

Although there is not enough evidence that today's hacker groups have strong technical experience in managing and manipulating AI-based IoT systems, they have probably already realized their enormous potential. Most of these criminal organizations are made up of hackers who are skilled at manipulating, exploiting and misusing any type of computer system. And this with attacks 24 hours a day, and from anywhere in the world (Velasco, 2022). With the use of AI-based BCI technologies, cybercriminals seem to have found a new vehicle to leverage their illegal activities and, in particular, new opportunities to design and carry out attacks against individuals, companies, and even governments. The literature has raised multiple hypotheses. Firstly, if a hacker takes control of BCIs connected to large number of people, he could manipulate them into voting for a particular candidate, a particular party or a particular issue, thereby secretly overthrowing a particular government and/ or entire infrastructures of a particular state. Although, for the moment, this seems like a highly fictitious scenario, the risk of certain hacker groups using BCIs to turn their "hosts" into a kind of army of programmable robots willing to do anything

[63] It is worth noting that both Interpol and Europol are aware of criminal activities carried out within the metaverse. In this regard, see Interpol. (2022, October 20). Interpol launches first global police Metaverse. https://clck.ru/3NkriG; Europol. (2022, October 21). Policing in the metaverse: what law enforcement needs to know. https://clck.ru/3Nkrnf

their "master" commands does not seem to be ruled out at all[64]. Although BCIs were designed by humans to hack the human brain, the same seems to happen with the risk of AI itself using BCIs to hack the human brain[65]. In fact, it seems that certain AI systems have the potential to become hackers themselves once they become «sentient» (Esmaeilzadeh & Vaezi, 2021)[66, 67]. If this is the case, everything indicates that they will have at their disposal the computerized means to assess the vulnerabilities of any type of social, economic, and political system, and then exploit them at an unprecedented speed, scale, and scope and in a way unimaginable by the limited human mind. It is not just a mere difference in level of intelligence. It is a difference of species. It may even happen that certain AI systems aim to crack other AI systems, with human beings themselves watching and constituting little more than mere collateral damage. Everything indicates that this scenario does not constitute hyperbole. In fact, none of these hypotheses require the creation of a science fiction technology from the distant future. It does not seem at all unreasonable to say that the development of AI will become so rapid that it will even surpass human understanding, as, in fact, it already seems to surpass it[68].

Conclusions

The aim of this study is to contribute to the study of neurohacking in the digital and artificial era and, above all, raise awareness about the neurosecurity (and also ethical) implications resulting from the malicious use of technologies associated with the metaverse and AI for neural manipulation purposes.

As a result of this study, it was found that the possible benefits of developing and (mass) implementing/ commercializing this type of technologies may not outweigh the possible disadvantages. As any computer, computer network, or most other forms of information communication technology (ICT), the hard core is based on electronic components with the capacity to process data, i.e. with the capacity to record, process, and store information (and information), and execute algorithms. As such, the overwhelming majority are, in principle, hackable. Rapid technological advances do not even constitute a deterrent. On the contrary, they end up becoming an attraction or even a challenge. Just think of the recent increase in cyberattacks on public organizations, or even cybercrime in general.

64 Lau, J. (2020, November 18). Hacking Humans: How Neuralink May Give AI The Keys To Our Brains. Forbes. https://clck.ru/3NkrqV

65 Schneier, B. (2021). The Coming AI Hackers. Belfer Center for Science and International Affairs. Harvard Kennedy School. https://clck.ru/3NkrrW

66 Ibid.

67 See also Johnson, A. (2024, March 19). Consciousness for Artificial Intelligence? IEEE Pulse. https://clck.ru/3Nkrtg

68 Schneier, B. (2021). The Coming AI Hackers. Belfer Center for Science and International Affairs. Harvard Kennedy School. https://clck.ru/3Nkrw7

While we are currently witnessing severe consequences in terms of financial and, often, reputational damage, in the near future, and in a world of a "multiverse" of metaverses with "smart technology", the consequences of implementing chips in the human brain to establish a kind of symbiosis with the internet and AI will be unimaginable. Although every detail of human life today is already monitored[69] and analyzed[70] by mobile devices (such as smartphones and/or smartwatches) in order to collect and store data for the purpose of creating detailed psychographic profiles, the human brain contains information that, by its very nature, cannot be recorded without neural interaction. The direct connection between the human brain, mobile devices and AI could establish a kind of access route to human consciousness. It could also allow certain groups of hackers (who are usually one step ahead of security protocols) to take control of the human mind, which includes the decision-making process and its execution. In this scenario, imagine the consequences in terms of voting and in the context of political elections, or even in the context of other more or less peaceful political movements. In the war and military sector, imagine the unlimited possibilities for creating "supersoldiers" or even in terms of surveillance and social control. Imagine, in a future limit and apocalyptic scenario, the eventuality that AI itself could take control of the neural interface and, thus, human consciousness.

In any case, it seems that, in the near future, it will be increasingly easy to plant ideas or even ideologies in people's heads (which nowadays mostly happens through social networks). To what extent will it be possible to guarantee the protection of personal data in the event of a "neuroattack"? The term "personal data" takes on a whole new level in this context. Combining the Freudian theory of the perception-consciousness system, to what extent is it possible to guarantee the protection of the "conscious psychic process" itself, including the previous unconscious state? Who does not remember the movie "Inception", starring Leonardo DiCaprio in 2010? In this context, there are even some philosophical questions. What is the real world? How can we know if it is real or not? Can life in the metaverse be considered "more real" than what is considered real life today?

Among Elon Musk's most famous warnings (often seen as a kind of savior of humanity[71]), two very peculiar ones stand out. In 2018, he said that AI could become an "immortal dictator"[72], from which humanity "will never be able to escape". In 2020,

---

69 Szoldra, P. (2016, September 16). This is everything Edward Snowden revealed in one year of unprecedented top-secret leaks. Business Insider. https://clck.ru/3Nkry8

70 Shane, S., Rosenberg, M., & Lehren, A. (2017, March 7). WikiLeaks Releases Trove of Alleged C.I.A. Hacking Documents. The New York Times. https://clck.ru/3Nkrzw

71 Dowd, M. (2017, March 26). Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse. The Vanity Fair. https://clck.ru/3NkkHp

72 Holley, P. (2018, April 6). Elon Musk's nightmarish warning: AI could become 'an immortal dictator from which we would never escape'. The Washington Post. https://clck.ru/3NkkGv

he warned again about the same issue, this time saying that "artificial intelligence will overtake humans in less than five years"[73]. Despite these, to say the least, intriguing warnings, the truth is that Elon Musk is also incessantly, obstinately and without apparent limits aiming to develop the functional and processing capacity of this type of technology, as well as expanding its possible forms of application. Take the current technological advances in robotics driven by the company "Tesla, Inc". Specifically, see the "Tesla Optimus" project[74] and the creation of humanoid robots nicknamed "Optimus" [75], whose appearance and characteristics are reminiscent of the robots in the movie "I, Robot", starring Will Smith in 2004, or even the Skynet robots in the movie "Terminator", starring Arnold Schwarzenegger in 1995. Are we unconsciously opening the doors to a fictional, dystopian world, like in the movie "The Matrix", starring Keanu Reeves in 1999? Only time will tell.

References

Abegão Alves, C. (2020). Contra a mente: ensaio de integração das lesões resultantes de intervenções na mente no artigo 143.º do Código Penal. In M. Fernanda Palma et al. (org.). *Livro em Memória do Professor Doutor João Curado Neves. Associação Académica da Faculdade de Direito de Lisboa* (pp. 215–235). AAFDL Editora. (In Portug.).

Bernal, S. L., Pérez, M. Q., Martínez Beltrán, E. T., Martínez Pérez, G., & Huertas Celdrán, A. (2022). When Brain-Computer Interfaces Meet the Metaverse: Landscape, Demonstrator, Trends, Challenges, and Concerns. *Computer Science – Human-Computer Interaction*. https://doi.org/10.48550/arXiv.2212.03169

Bublitz, J. C., & Merkel, R. (2014). Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. *Criminal Law and Philosophy*, *8*, 51–77. https://doi.org/10.1007/s11572-012-9172-y

Czech, A. (2021). Brain-Computer Interface Use to Control Military Weapons and Tools. In S. Paszkiel (Ed.), *Control, Computer Engineering and Neuroscience. ICBCI 2021. Advances in Intelligent Systems and Computing* (Vol. 1362). Springer, Cham. https://doi.org/10.1007/978-3-030-72254-8_20

Denning, T., Matsuoka, Y., & Kohno, T. (2009). Neurosecurity: Security and privacy for neural devices. *Neurosurgical Focus*, *27*(1), E7. https://doi.org/10.3171/2009.4.focus0985

Dias Venâncio, P. (2011). *Lei do Cibercrime. Anotada e Comentada*. Almedina. (In Portug.).

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, Bh., Buhalis, D., Cheung, Ch. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, Ch., Jebabli, I., Janssen, … Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *66*, 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Ford, M. (2016). *Robôs: A Ameaça de um futuro sem emprego*. Bertrand Editora. (In Portug.).

Esmaeilzadeh, H., & Vaezi, R. (2021). Conscious AI. *arXiv:2105.07879*. https://doi.org/10.48550/arXiv.2105.07879

Ienca, M., & Haselager, P. (2016). Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, *18*, 117–129. https://doi.org/10.1007/s10676-016-9398-9

[73] Cuthbertson, A. (2020, July 27). Elon Musk claims AI will overtake humans 'in less than five years'. Independent. https://clck.ru/3NkkF2

[74] Levin, T. (2022, January 27). Elon Musk says Tesla's humanoid robot is the most important product it's working on – and could eventually outgrow its car business. Business Insider. https://clck.ru/3NkkDD

[75] Gomez, B. (2021, August 24). Elon Musk warned of a 'Terminator'-like AI apocalypse – now he's building a Tesla robot. CNBC. https://clck.ru/3NkkBk

Kotchetkov, I., Hwang, B. Y., Appelboom, G., Kellner, Ch. P., & Connolly E. S. Jr. (2010). Brain-computer interfaces: Military, neurosurgical, and ethical perspective. *Neurosurgical Focus*, *28*(5), E25. https://doi.org/10.3171/2010.2.focus1027

Li, Q., Ding, D., & Conti, M. (2015). Brain-computer interface applications: Security and privacy challenges. *IEEE conference on communications and network security (CNS). Florence, 2015* (pp. 663–666). https://doi.org/10.1109/CNS.2015.7346884

Marques, G., & Martins, L. (2006). *Direito da informática* (2.ª Edição). Almedina. (In Portug.).

Miranda, R., Casebeer, W. D., Hein, A. M., Judy, J. W., Krotkov, E. P., Laabs, T. L., Manzo, J. E., Pankratz, K. G., Pratt, G. A., Sanchez, J. C., Weber, D. J., Wheeler, T. L., & Ling, G. S. F. (2015). DARPA-funded efforts in the development of novel brain-computer interface technologies. *Journal of Neuroscience Methods*, *244*, 52–67. https://doi.org/10.1016/j.jneumeth.2014.07.019.

Mochan, A., Parkin, B., Farinha, J., & Bailey, G. (2025). *Emerging applications of neurotechnology and their implications for EU governance*. Publications Office of the European Union. Luxembourg.

Moulin, T. (2022). 'I Will Control Your Mind': The International Regulation of Brain-Hacking. *San Diego International Law Journal*, *24*(65).

Orlosky, J., Sra, M., Bektaş, K., Peng, H., Kim, J., Kosmyna, N., Höllerer, T., Steed, A., Kiyokawa, K., Akşit, K. (2021). Telelife: The Future of Remote Living. *Frontiers in Virtual Reality*, *2*, 763340. https://doi.org/10.3389/frvir.2021.763340

Pooyandeh, M., Han, K. J., & Sohn, I. (2022). Cybersecurity in the AI-Based Metaverse: A Survey. *Applied Sciences. MDPI*, *12*(24), 12993. https://doi.org/10.3390/app122412993

Radu, R. (2024). Neurotechnologies and the future of internet governance. Technical Report. EUI. RSC (Global Governance Programme). *GIFI. Issue Brief*.

Rodrigues, B. S. (2009). *Direito Penal Especial. Direito Penal Informático-Digital*. Almedina. (In Portug.).

Rosenfeld, P., Biroschak, J., & Furedy, J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology*, *60*(3), 251–259. https://doi.org/10.1016/j.ijpsycho.2005.06.002

Rosenfeld, P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds), *Memory detection: Theory and application of the concealed information test* (pp. 63–89). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.005

Shen, F. (2013). Mind, Body, and the Criminal Law. *Minnesota Law Review*, *97,* 2036–2175.

Vallabhaneni, A., Wang, T., & He, B. (2005). Brain – Computer Interface. In B. He (Ed.), *Neural Engineering. Bioelectric Engineering* (pp. 85–121). Springer. Boston, MA. https://doi.org/10.1007/0-306-48610-5_3

Van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., & Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, *6*(4), 041001. https://doi.org/10.1088/1741-2560/6/4/041001

Vasconcelos Casimiro, S. (2000). *A responsabilidade civil pelo conteúdo da informação transmitida pela Internet*. Almedina. (In Portug.).

Velasco, C. (2022). Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments. *ERA Forum*, *23*, 109–126. https://doi.org/10.1007/s12027-022-00702-z

Yuan, B., Hsieh, Chih-Hung, & Chang, Chien-Ching (2010). National technology foresight research: A literature review from 1984 to 2005. *International Journal of Foresight and Innovation Policy*, *6*(1), 5–35. http://dx.doi.org/10.1504/IJFIP.2010.032663

Zhang, X., Ma, Z., Zheng, H., Li, T., Chen, K., Wang, X., Liu, Ch., Xu, L., Wu, X., Lin, D., & Lin, H. (2020). The combination of brain-computer interfaces and artificial intelligence: applications and challenges. *Annals of Translational Medicine*, *8*(11), PMC7327323. https://doi.org/10.21037/atm.2019.11.109

Author information

Diogo P. Coelho – PhD student, University of Seville
Address: 4 Calle San Fernando, 41013 Sevilla, Spain
E-mail: diopercoe@alum.us.es
ORCID ID: https://orcid.org/0000-0002-2082-1231
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57703490300
WoS Researcher ID: https://www.webofscience.com/wos/author/record/GLU-8923-2022
Google Scholar ID: https://scholar.google.com/citations?user=-laUdL8AAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – May 5, 2025
Date of approval – May 26, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:343.721:004.8
EDN: https://elibrary.ru/smgmxq
DOI: https://doi.org/10.21202/jdtl.2025.16

Нейрохакинг в эпоху цифровизации и искусственного интеллекта: правовые аспекты защиты нейронной информации

Диого Перейра Коэльо

Севильский университет, Севилья, Испания

Ключевые слова

интерфейс «мозг – компьютер»,
искусственный интеллект,
киберпреступность,
метавселенная,
нейробезопасность,
нейропреступность,
нейротехнологии,
нейрохакинг,
право,
цифровые технологии

Аннотация

Цель: внесение вклада в осмысление концепции нейропреступности, а также в изучение текущих и будущих рисков с точки зрения нейробезопасности в условиях развития цифровизации и искусственного интеллекта.

Методы: в исследовании применен критико-описательный анализ связи между киберпреступностью и нейропреступностью, проведено концептуальное разграничение интерфейса «мозг – компьютер» и вариантов его использования, выполнено описание различий между нейронными и психическими манипуляциями. Исследуется правовая автономия преступлений против психической неприкосновенности по отношению к преступлениям против физической неприкосновенности. Методологический аппарат включает анализ существующих прототипов нейропреступлений на основе четырехфазного цикла интерфейса «мозг – компьютер» и изучение специфики нейрохакинга в контексте метавселенной и технологий искусственного интеллекта.

Результаты: исследование выявило сущностные характеристики нейрохакинга как неправомерного использования нейронных устройств для получения несанкционированного доступа к нейронной информации и ее манипулирования. Определены четыре основных типа приложений интерфейса «мозг – компьютер», подверженных нейрохакингу: нейромедицинские приложения, системы аутентификации пользователей, видеоигры и приложения на базе смартфонов. Установлены модальности нейрохакинга на каждой фазе цикла интерфейса «мозг – компьютер»: манипуляции на этапе ввода нейронной информации, измерения и записи мозговой активности, декодирования и классификации нейронной информации, а также на этапе вывода результата. Проанализированы специфические угрозы нейрохакинга в эпоху цифровизации, включая иммерсивные атаки и атаки типа «человек – джойстик» в метавселенной.

Научная новизна: впервые проведено комплексное разграничение концепций нейропреступности и киберпреступности с выделением их специфических правовых последствий. Предложена авторская классификация нейропреступлений на основе четырехфазного цикла интерфейса «мозг – компьютер». Обоснована необходимость выделения психической неприкосновенности как самостоятельного объекта правовой защиты, отличного от защиты физической неприкосновенности. Впервые исследованы особенности нейрохакинга в контексте метавселенной и технологий искусственного интеллекта, включая анализ новых типов атак и угроз нейробезопасности.

Практическая значимость: результаты исследования имеют важное значение для развития правового регулирования в области нейробезопасности и разработки соответствующих нормативных актов. Выявленные типы нейропреступлений и их классификация могут служить основой для создания специализированного законодательства о защите нейронных данных и психической неприкосновенности. Практические рекомендации по обеспечению нейробезопасности интерфейсов «мозг – компьютер» востребованы в медицинской практике, индустрии видеоигр, системах аутентификации и разработке приложений для смартфонов.

Для цитирования

Коэльо, Д. П. (2025). Нейрохакинг в эпоху цифровизации и искусственного интеллекта: правовые аспекты защиты нейронной информации. *Journal of Digital Technologies and Law*, *3*(3), 397–430. https://doi.org/10.21202/jdtl.2025.16

Список литературы

Abegão Alves, C. (2020). Contra a mente: ensaio de integração das lesões resultantes de intervenções na mente no artigo 143.º do Código Penal. In M. Fernanda Palma et al. (org.). *Livro em Memória do Professor Doutor João Curado Neves. Associação Académica da Faculdade de Direito de Lisboa* (pp. 215–235). AAFDL Editora. (In Portug.).

Bernal, S. L., Pérez, M. Q., Martínez Beltrán, E. T., Martínez Pérez, G., & Huertas Celdrán, A. (2022). When Brain-Computer Interfaces Meet the Metaverse: Landscape, Demonstrator, Trends, Challenges, and Concerns. *Computer Science – Human-Computer Interaction*. https://doi.org/10.48550/arXiv.2212.03169

Bublitz, J. C., & Merkel, R. (2014). Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. *Criminal Law and Philosophy*, *8*, 51–77. https://doi.org/10.1007/s11572-012-9172-y

Czech, A. (2021). Brain-Computer Interface Use to Control Military Weapons and Tools. In S. Paszkiel (Ed.), *Control, Computer Engineering and Neuroscience. ICBCI 2021. Advances in Intelligent Systems and Computing* (Vol. 1362). Springer, Cham. https://doi.org/10.1007/978-3-030-72254-8_20

Denning, T., Matsuoka, Y., & Kohno, T. (2009). Neurosecurity: Security and privacy for neural devices. *Neurosurgical Focus*, *27*(1), E7. https://doi.org/10.3171/2009.4.focus0985

Dias Venâncio, P. (2011). *Lei do Cibercrime. Anotada e Comentada*. Almedina. (In Portug.).

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, Bh., Buhalis, D., Cheung, Ch. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, Ch., Jebabli, I., Janssen, … Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *66*, 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Ford, M. (2016). *Robôs: A Ameaça de um futuro sem emprego*. Bertrand Editora. (In Portug.).

Esmaeilzadeh, H., & Vaezi, R. (2021). Conscious AI. *arXiv:2105.07879*. https://doi.org/10.48550/arXiv.2105.07879

Ienca, M., & Haselager, P. (2016). Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, *18*, 117–129. https://doi.org/10.1007/s10676-016-9398-9

Kotchetkov, I., Hwang, B. Y., Appelboom, G., Kellner, Ch. P., & Connolly E. S. Jr. (2010). Brain-computer interfaces: Military, neurosurgical, and ethical perspective. *Neurosurgical Focus*, *28*(5), E25. https://doi.org/10.3171/2010.2.focus1027

Li, Q., Ding, D., & Conti, M. (2015). Brain-computer interface applications: Security and privacy challenges. *IEEE conference on communications and network security (CNS). Florence, 2015* (pp. 663–666). https://doi.org/10.1109/CNS.2015.7346884

Marques, G., & Martins, L. (2006). *Direito da informática* (2.ª Edição). Almedina. (In Portug.).

Miranda, R., Casebeer, W. D., Hein, A. M., Judy, J. W., Krotkov, E. P., Laabs, T. L., Manzo, J. E., Pankratz, K. G., Pratt, G. A., Sanchez, J. C., Weber, D. J., Wheeler, T. L., & Ling, G. S. F. (2015). DARPA-funded efforts in the development of novel brain-computer interface technologies. *Journal of Neuroscience Methods*, *244*, 52–67. https://doi.org/10.1016/j.jneumeth.2014.07.019.

Mochan, A., Parkin, B., Farinha, J., & Bailey, G. (2025). *Emerging applications of neurotechnology and their implications for EU governance*. Publications Office of the European Union. Luxembourg.

Moulin, T. (2022). 'I Will Control Your Mind': The International Regulation of Brain-Hacking. *San Diego International Law Journal*, *24*(65).

Orlosky, J., Sra, M., Bektaş, K., Peng, H., Kim, J., Kosmyna, N., Höllerer, T., Steed, A., Kiyokawa, K., Akşit, K. (2021). Telelife: The Future of Remote Living. *Frontiers in Virtual Reality*, *2*, 763340. https://doi.org/10.3389/frvir.2021.763340

Pooyandeh, M., Han, K. J., & Sohn, I. (2022). Cybersecurity in the AI-Based Metaverse: A Survey. *Applied Sciences. MDPI*, *12*(24), 12993. https://doi.org/10.3390/app122412993

Radu, R. (2024). Neurotechnologies and the future of internet governance. Technical Report. EUI. RSC (Global Governance Programme). *GIFI. Issue Brief*.

Rodrigues, B. S. (2009). *Direito Penal Especial. Direito Penal Informático-Digital*. Almedina. (In Portug.).

Rosenfeld, P., Biroschak, J., & Furedy, J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology*, *60*(3), 251–259. https://doi.org/10.1016/j.ijpsycho.2005.06.002

Rosenfeld, P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds), *Memory detection: Theory and application of the concealed information test* (pp. 63–89). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.005

Shen, F. (2013). Mind, Body, and the Criminal Law. *Minnesota Law Review*, *97*, 2036–2175.

Vallabhaneni, A., Wang, T., & He, B. (2005). Brain – Computer Interface. In B. He (Ed.), *Neural Engineering. Bioelectric Engineering* (pp. 85–121). Springer. Boston, MA. https://doi.org/10.1007/0-306-48610-5_3

Van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., & Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, *6*(4), 041001. https://doi.org/10.1088/1741-2560/6/4/041001

Vasconcelos Casimiro, S. (2000). *A responsabilidade civil pelo conteúdo da informação transmitida pela Internet*. Almedina. (In Portug.).

Velasco, C. (2022). Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments. *ERA Forum*, *23*, 109–126. https://doi.org/10.1007/s12027-022-00702-z

Yuan, B., Hsieh, Chih-Hung, & Chang, Chien-Ching (2010). National technology foresight research: A literature review from 1984 to 2005. *International Journal of Foresight and Innovation Policy*, *6*(1), 5–35. http://dx.doi.org/10.1504/IJFIP.2010.032663

Zhang, X., Ma, Z., Zheng, H., Li, T., Chen, K., Wang, X., Liu, Ch., Xu, L., Wu, X., Lin, D., & Lin, H. (2020). The combination of brain-computer interfaces and artificial intelligence: applications and challenges. *Annals of Translational Medicine*, *8*(11), PMC7327323. https://doi.org/10.21037/atm.2019.11.109

Сведения об авторе

Коэльо Диого Перейра – аспирант, Севильский университет
Адрес: Испания, 41013, г. Севилья, Калле Сан Фернандо, д. 4
E-mail: diopercoe@alum.us.es
ORCID ID: https://orcid.org/0000-0002-2082-1231
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57703490300
WoS Researcher ID: https://www.webofscience.com/wos/author/record/GLU-8923-2022
Google Scholar ID: https://scholar.google.com/citations?user=-laUdL8AAAAJ

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс
Специальность ВАК: 5.1.4 / Уголовно-правовые науки

История статьи

Дата поступления – 5 мая 2025 г.
Дата одобрения после рецензирования – 26 мая 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.

Research article
UDC 34:004:340.1:004.8
EDN: https://elibrary.ru/pltfwo
DOI: https://doi.org/10.21202/jdtl.2025.17

Agentic Artificial Intelligence: Legal and Ethical Challenges of Autonomous Systems

Gordon Bowen

Anglia Ruskin University, Cambridge, United Kingdom

Keywords

agentic artificial intelligence,
artificial intelligence,
autonomy,
digital technologies,
ethics,
law,
legal regulation,
liability,
programming,
risk

Abstract

Objective: to identify specific legal and ethical problems of agentic artificial intelligence and develop recommendations for the creation of protective mechanisms to ensure the responsible functioning of autonomous AI systems.

Methods: the research is conceptual in nature and is based on a systematic analysis of scientific literature on the ethics of artificial intelligence, legal regulation of autonomous systems and social interaction of AI agents. The work uses a comparative analysis of various types of AI systems, a study of the potential risks and benefits of agentic artificial intelligence, as well as an interdisciplinary approach that integrates advances in law, ethics, and computer science to form a comprehensive understanding of the issue.

Results: the research has established that agentic artificial intelligence, possessing the decision-making autonomy and ability to social interaction, creates qualitatively new legal and ethical challenges compared to traditional AI assistants. The main categories of potential harm were identified: direct impact on users through overt and covert actions, manipulative influence on behavior, and cumulative harm from prolonged interaction. The author stipulates the need for distributing responsibility between three key actors: the user, the developer and the owner of the agentic artificial intelligence system.

Scientific novelty: for the first time, the research presents a systematic analysis of the ethical aspects of agentic artificial intelligence as a qualitatively new class of autonomous systems that differ from traditional AI assistants in the degree of independence and social interactivity. The author developed a typology of potential risks of social interaction with agent-based intelligent systems and proposes a conceptual model for the distribution of legal and ethical responsibilities in the user-developer-owner triad.

Practical significance: the research forms the theoretical basis for the development of ethical principles and legal norms governing agentic based artificial intelligence in a growing market for autonomous intelligent systems. The findings will be useful for legislators creating a regulatory framework, developers designing protective mechanisms, as well as organizations implementing agentic artificial intelligence systems in various economic fields.

For citation

Bowen, G. (2025). Agentic Artificial Intelligence: Legal and Ethical Challenges of Autonomous Systems. *Journal of Digital Technologies and Law, 3*(3), 431–445. https://doi.org/10.21202/jdtl.2025.17

Contents

Introduction

The capabilities of AI agents, such as communications skills and sophisticated reasoning that does not require human intervention, are increasing. AI assistants are anchored to users, but agentic AI agents have degrees of freedom[1]. The global market value of agentic AI in 2024 was US$5.1 billion, which is expected to increase to US$47 billion by 2030, with a compound annual growth rate of 44 % [2]. The degree of freedom now experienced by agentic AI agents re- quires legal and ethical frameworks to regulate agentic AI agents' behaviour. The owner/developer of an agentic AI agent and an agentic AI agent's software behaviour both need monitoring from a legal and ethical perspective. However, a balance needs to be struck to not lose the advantages of agentic AI agents. The ethics and legal frameworks for "static" AI assistants, which are controlled or tethered to ethical behaviour, require redefining for agentic AI agents. How should these frameworks be changed for agentic AI agents? Do agentic AI agents require social consciousness

1 Morris, B. (2024). Beyond Intelligence: The Impact of Advanced AI Agents. https://clck.ru/3NedB2

2 Vailshery, L. S. (2025). Global market value of agentic AI 2030. https://clck.ru/3NedDe

to navigate the new ethics and legal environments? The paper is structured as follows: introduction, literature review, implications and conclusion. The overarching aim of the paper is to understand how the debate on ethics and the legal landscape for agentic AI agents will need to evolve.

1. Literature review

1.1. Three types of AI agents

AI agents are also known as compound AI systems and are a growing area of research (Kapoor et al., 2024). Compound AI systems are the best way to leverage and maximise AI models and may have been one of the most important trends in 2024[3]. Compound AI systems differ from AI systems (large language models) in many ways, for example, they tackle and complete harder tasks, they have more real-world use and can solve problems that do not have a single answer; compound AI systems might require custom-built agent–computer interfaces (Yang, Jimenez et al., 2024). A compound AI system can manage several agents, which has a price implication (Kapoor et al., 2024).

In traditional AI, agents were considered to be able to perceive and act upon the environment (Russell & Norvig, 1995); from a traditional perspective, a thermostat could be classified as an agent (Kapoor et al., 2024). Agentic AI systems are often viewed as a spectrum of AI systems with more or less agentic capability[4]. There are three types of AI agents (Alberts, Keeling et al., 2024): Artifacts (an agentic agent interprets data in a social environment), Interactive Systems (behave as a social actor) and Conversational Agents (the agents have social roles). AI agents as social actors need to be conversational and thus interact with the user, but this goes beyond being agreeable, friendly, truthful and not using harmful language. Interactions are expected to be contextual, which requires awareness of the individual user, the social environment and the situational context. This will translate into AI agents giving information unprompted, which leads to them making suggestions (Alberts, Keeling et al., 2024).

1.2. The social interaction of AI agents

Technologies can be considered social because they are situated and embedded in social environments. This assumption is accepted by researchers who research the ideological perspective of technology, such as culture biases and values that are embedded in the technologies we use (Bender et al., 2021; Shelby et al., 2023). Systems that are not socially interactive can be seen to be harmful (Alberts, Keeling et al., 2024). An example

[3] Zaharia, M., Khattab, O., Chen, L., Davis, J. Q., Miller, H., Potts, C., Zou, J., Carbin, M., Frankle, J., Rao, N., & Ghodsi, A. (2024). The Shift from Models to Compound AI Systems. https://clck.ru/3NedKd

[4] Ng, A. (2024). Welcoming Diverse Approaches Keeps Machine Learning Strong. https://clck.ru/3NedNZ

of harm in a "passive" technological system includes training data that misrepresent demographics (Bender et al., 2021).

Looking beyond passive technologies, people treat interactive technology as having a purpose/intention and social meaning (Grimes et al., 2021). Furthermore, the situated actions of a system are interpreted in human social ways, which is a central tenet of the Computers Are Social Actors research philosophy, where humans interacting with computer systems apply human social norms and expectations to their interactions with technology (Nass et al., 1994).

Interactive systems mimic human behaviour or qualities. This is achieved through social cues (speaking in the first person and expressing emotion) (Grimes et al., 2021). Conversational agents can respond in familiar social situations. An AI system could be situated in a social role in which the AI agent is a friend or therapist; however, the AI agent could utter something offensive that could upset the user, or familiarity in a relationship with the user could breed contempt and insensitivity (Alberts, Keeling et al., 2024). Social interactions that could cause harm are categorised as follows (Alberts, Keeling et al., 2024):

– Direct harm to the user – giving rise to overt action such as offensiveness due to the language used or behaviour.

– Direct harm to the user – giving rise to covert action such as opinions that appear positive or neutral.

– Interactions that exert a harmful influence on behaviour – misleading or giving false information.

– Interactions that exert a harmful influence on behaviour – manipulating and persuading users to do things they would not normally do.

– Interactions that collectively harm users – harm emerging over time in relationships.

Harm from interactions means that harm resides inside language (Shelby et al., 2023). Direct harm to the user is dependent on the contextual language and relational language. Language can be conceived as positive (endearment using derogatory language) or less positive (women or elderly being patronised) depending on the situation (Coghlan et al., 2021).

Interactions from an influencer cause secondary harm effects by influencing the thinking or doing of an individual. Thus, agentic AI can have undue influence on an individual by producing false or misleading behaviour in the individual (Alberts, Keeling et al., 2024). Engaging in social cues makes systems more intuitive and engaging (Kocielnik et al., 2021); humans react to social cues emotionally and not rationally, and this can be used to manipulate individuals (Alberts, Lyngs et al., 2024; Shamsudhin & Jotterand, 2021).

Interactions that collectively cause harm include dismissive actions that are tactless and controlling. The collective effect of harm is cumulative, for example, a single instance of tactlessness can be ignored, but if it is repeated then it could affect an individual adversely over time (Alberts, Keeling et al., 2024).

1.3. Agentic AI decision making

Agentic AI systems require unprecedented autonomy and contextual awareness (Martinez & Kifle, 2024; Mohanarangan et al., 2024). The decision-making process in the agentic AI algorithm needs to be revolutionary to fulfil the requirements to work independently and make logical and coherent decisions in the environment it operates within. The algorithm will be making real-time decisions and synthesising complex data and datasets (Abuelsaad et al., 2024). Agentic AI has two capabilities that go beyond the capabilities of AI assistants. The first is decision making that operates at different levels, from low-level responses to high-level strategic responses, which requires long-term thoughtfulness (Abuelsaad et al., 2024). A second capability is the move from reactive to proactive goal-oriented behaviour, which requires the system to identify complex tasks and determine the necessary subtasks. Thus, to pursue its objectives, agentic AI will require a flexible architecture in its goal management software (Martinez & Kifle, 2024). Agentic AI requires an adaptive learning style that can harness different learning styles, and the ability to reinforce learning so it can proactively apply the learning style that is appropriate to situations. The adaptive learning system must be able to learn from experience (Abuelsaad et al., 2024).

Agentic AI acts as an outsourcer for an organisation. Early adopters will have a first mover advantage (market position, innovation, customer relations, operational efficiencies, learning curve, market share), and last movers will potentially lose their competitive advantage (incur a loss of market share and increased costs, experience slowness in business and process innovation, lag in personalisation of customer services, experience higher opportunity costs leading to higher operational costs, miss early learning opportunities, and potentially experience a higher barrier to entry and pay a lower entry fee to enter the market but with fewer tests) (Beulen et al., 2022). Agentic AI offers many benefits: positive impacts on operating costs; higher efficiency because AI can perform tasks automatically and with greater accuracy; scalability without the need for additional resources and investment; and core goal focus, because organisations can focus on their core business activities and leave the minor or less important activities to AI (Hosseini & Seilani, 2025). However, the use of agentic AI has drawbacks: dependence on technology (overdependence

on technology could lead to operational disruption in AI service); limited range of personalisation because many of the tasks require extensive customisation; privacy and security issues, for example, the outsourcing of data to third parties raises concerns on privacy and security; and hidden costs relating to training, deployment and implementation.

Applications of agentic AI will span many industries, including robotics and manufacturing[5], healthcare systems[6], transport and logistics[7], traffic management systems[8] and financial services[9]. One emerging application of agentic AI is dynamic patient needs, which leads to personalised medicines (Hasan et al., 2025). In this application, agentic AI manages patients with chronic illnesses by overseeing patient history and sending reminders to patients (Yang, Garcia et al., 2024); this leads to recommendations on treatment using health indicators. This type of agentic AI system could manage individualistic patient care management and monitor for early warning signs of the health progression of patients, especially for older patients (Acharya et al., 2025). Another application of agentic AI is the automatic generation of new content that targets wider audiences and meets content requirements based on set criteria. This application would be helpful to marketing activities, such as sending customised emails to customers and potential customers. Literature searches by businesses, scientists and academics would be faster with agentic AI, and lead to new thoughts and ideas. Agentic AI could empower drug discovery, development and delivery (Gao et al., 2024).

Agentic AI research is gaining traction in moral reasoning and ethical decision making (Small & Lew, 2021). The importance of privacy and security in the handling of sensitive information has pushed this type of research to the fore. Research on moral reasoning seeks to establish an ethical basis for autonomous systems so that agentic AI systems can select actions with thought of their effects and values. Thus, in this context, the integration of psychology, ethics and philosophy create an overarching goal for AI systems that is ethical. All agentic systems need to be ethical in their decision making, especially health, autonomous, and law and order systems, because these decisions influence society (Acharya et al., 2025).

---

5 Randieri, C. (2025, January 3). Agentic AI: A New Paradigm In Autonomous Artificial Intelligence. Forbes. https://clck.ru/3NedZf

6 Automation Anywhere. (n.d.). What is agentic AI? Key benefits & features. https://clck.ru/3Nedsx

7 Ibid.

8 Randieri, C. (2025, January 3). Agentic AI: A New Paradigm In Autonomous Artificial Intelligence. Forbes. https://clck.ru/3NedZf

9 Ibid; Automation Anywhere. (n.d.). What is agentic AI? Key benefits & features. https://clck.ru/3Nedsx

Agentic AI systems require self-awareness and meta-cognition (Langdon et al., 2022), which can be achieved by building systems that understand their actions, abilities and limitations as self-referential knowledge. Self-awareness in AI systems can be done through self-evaluation on whether they have carried out tasks optimally, what can be improved, and what actions should be taken when failures occur or performance is poor. Self-agency skills (carrying out of the task) and ability (to detect the need to carry out the task) will enable agentic AI to assess its strategies and learning processes to improve the effectiveness of its decision making. Progress in research on self-awareness and meta-cognition might lead to more flexible and sophisticated agentic AI systems, thereby leading to enhanced and improved performance with the robustness to operate in multi-environments (Acharya et al., 2025). This will require further research on creating AI agency models, adaptive moral frameworks and contextual decision making (Lai et al., 2021).

2. Implications

Compound AI agents are more powerful than AI systems; thus, the ethical and legal issues are more complex. This is exacerbated by the social independence of agentic AI agents. A user/owner/developer has a degree of control over passive AI systems (AI assistants) because passive AI systems are tethered to a position and are singular in problem solving or tasks.

The owner/developer of an AI agent, the user of agentic AI, and agentic AI agent algorithm all need to behave ethically. An ethical and legal perspective of an agentic agent's interactions with a user needs to be taken so that the agentic agent behaves responsibly and does not cause harm. The technical developer and the owner of the algorithm need to ensure that an agentic AI system is applied ethically and legally. Why? Agentic software becomes independent once released; thus, guardrails need to be in place to monitor its behaviour and personality. Where does the legal responsibility lie if an agentic AI system runs amok or goes rogue? What if it causes harm, for example in the retrieving of data and information from a third party because it has a degree of decision-making capability, and one could argue consciousness[10] (Lim et al., 2025). However, the user of the AI agentic algorithm might have some ethical and legal responsibilities. What if a user asked an agentic AI system to do something that is unethical and illegal, such as transferring information without due process? In this situation, who is liable? Is it the user or the developer/owner of the AI algorithm? What if the relationship between an agentic AI and a user becomes toxic, and the agentic AI goes rogue and causes harm

[10] Al-Sibai, N. (2022). OpenAI Chief Scientist Says Advanced AI May Already Be Conscious. https://clck.ru/3Nee2Z

(Alberts, Keeling et al., 2024)? Agentic AI can contextualise the environmental landscape, so they have a sense of awareness, but does this let the user off the hook? There are similarities between agentic AI and autonomous automobiles. Parties will try to absolve themselves of blame and liability.

The problems and issues identified are not that prevalent with AI assistants. Moving from passive technological systems to interactive systems raises additional concerns not only about the legal and ethical implications but also about the scope and ability of agentic AI systems. Agentic AI is the future direction of AI, and it is becoming unstoppable, given the many benefits; however, there are challenges which need to be acknowledged and acted upon by the AI community for the protection of society and humankind. Treating individuals with respect is the starting point to making agentic AI ethically and legally responsible. Basic Psychological Needs Theory and the field of human–robot interaction could assist in the development of suitable frameworks (Li et al., 2025; Hosseini & Seilani, 2025; Korzynski et al., 2025; Kshteri, 2025).

New applications of agentic AI have been established in healthcare, logistics and transportation, and financial services. However, security and privacy concerns are rife because of the level of data and the independence of actions in the decision making of agentic AI. It makes decisions by breaking down complex tasks and compartmentalising them into subunits. The question is: How robust is the decision-making architecture and the understanding of the environmental ecosystem in which the decision-making process operates? Reliability and accuracy of the decisions is underpinned by, and dependent on, these areas. The decision-making process and the environmental ecosystem are starting points and foundational for the decision outcomes. An unreliability of the foundational aspects of agentic AI could contribute to the algorithm running amok and suffering from hallucinations. The range of applications and emerging applications of agentic AI makes it necessary for guardrails to be implemented at all levels of the architecture, which requires a hierarchical architecture of the agentic AI system. However, feedback will be needed to test subsystems of the various architectures of the algorithm so parts that are underperforming or are exhibiting worrying behaviour can be isolated or corrected. Will this require redundancy in the agentic AI architecture? If this is the case, then costs to own and implement agentic systems will rise. The application of the reasoning and consciousness that are speculated to exist in AI systems could be a pointer in the right direction. The emailing of customers and potential customers by agentic AI has business-related risks; thus, guardrails are necessary throughout the agentic AI system. Business-affecting problems risk reputational damage, brand credibility and relationship damage. The benefits of applying agentic AI in new and emerging applications, such as drug development, could drive the application without the necessary guardrails and reliability in the architecture in place. Does the societal value of agentic AI outweigh the benefits of ensuring rigorousness in safeguarding the regulatory and legal frameworks? Should agentic AI safeguarding, legal and regulatory, be more trial and error or experiential learning?

Conclusions

Compound AI agents (agentic AI) have many benefits that range from being able to work independently to the ability to reason; hence, they have a level of awareness and consciousness. Nevertheless, there are dangers, which requires a balanced approached to the deployment of agentic AI. Autonomous automobile scenarios are applicable to agentic AI, and lessons learnt from the autonomous automobile industry is a good starting point to understand the ethical and legal situations that are applicable to agentic AI. The risks of agentic AI need to be balanced with suitable guardrails that do not reduce or hinder innovation in the application of agentic AI. This requires an evolving legal and ethical framework that continues to protect society but also delivers the benefits of agentic AI to businesses and industries. Agentic AI takes decision making from human–machine interface to machine–machine interaction without the need for human intervention in decisions, but what if something goes wrong. Guardrails need to be implemented that are rigorous, resilient and robust to sustain ethical and legal frameworks.

References

Abuelsaad, T., Akkil, D., Dey, P., Jagmohan, A., & Vempaty, A. (2024). Agent-E: From Autonomous Web Navigation to Foundational Design Principles in Agentic Systems. *arXiv preprint arXiv:2407.13032*. https://doi.org/10.48550/arXiv.2407.13032

Acharya, D. B., Kuppan, K., & Ashwin, D. B. (2025). Agentic AI: Autonomous intelligence for complex goals – a comprehensive survey. In *IEEE Access (*vol. 13, pp. 18912–18936). https://doi.org/10.1109/ACCESS.2025.3532853

Alberts, L., Keeling, G., & McCroskery, A. (2024). Should agentic conversational AI change how we think about ethics? Characterising an interactional ethics centred on respect. *arXiv:2401.09082v2*. https://doi.org/10.48550/arXiv.2401.09082

Alberts, L., Lyngs, U., & Van Kleek, M. (2024). Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–25. https://doi.org/10.1145/3653693

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922

Beulen, E., Plugge, A., & van Hillegersberg. J. (2022). Formal and relational governance of artificial intelligence outsourcing. *Information System E Business Management*, *20*(4), 719–748. https://doi.org/10.1007/s10257-022-00562-7

Coghlan, S., Waycott, J., Lazar, A., & Neves, B. (2021). Dignity, Autonomy, and Style of Company: Dimensions Older Adults Consider for Robot Companions. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449178

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J., & Zitnik, M. (2024). Empowering biomedical discovery with AI agents. *Cell*, *187*(22), 6125–6151. https://doi.org/10.1016/j.cell.2024.09.022

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515.

Hasan, S. S., Fury, M. S., Woo, J. J., Kunze, K. N., & Ramkumar, P. N. (2025). Ethical Application of Generative Artificial Intelligence in Medicine. *Arthroscopy: Journal of Arthroscopic Related Surgery*, *41*(4), 874–885. https://doi.org/10.1016/j.arthro.2024.12.011

Hosseini, S., & Seilani, H. (2025). The Role of Agentic AI in Shaping a Smart Future: A Systematic review. *Array*, *26*, 100399. https://doi.org/10.1016/j.array.2025.100399

Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2024). AI Agents That Matter. *arXiv:2407.01502v1*.

Kocielnik, R., Langevin, R., George, J. S., Akenaga, S., Wang, A., Jones, D. P., Argyle, A., Fockele, C., Anderson, L., Hsieh, D. T., Kabir, Y., Duber, H., Hsieh, G., & Hartzler, A. L. (2021). Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain*. ACM, New York, NY, USA. https://doi.org/10.1145/3469595.3469599

Korzynski, P., Edwards, A., Gupta, M. C., Mazurek, G., & Wirtz, J. (2025). Humanoid robotics and agentic AI: reframing management theories and future research directions. *European Management Journa*l, *43*(4), 548–560. https://doi.org/10.1016/j.emj.2025.06.002

Kshetri, N. (2025). Transforming cybersecurity with agentic AI to combat emerging cyber threats. *Telecommunications Policy*, *49*(6), 102976. https://doi.org/10.1016/j.telpol.2025.102976

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a science of human-AI decision making: A survey of empirical studies. *arXiv:2112.11471*. https://doi.org/10.48550/arXiv.2112.11471

Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Network*, *145*, 80–89. https://doi.org/10.1016/j.neunet.2021.10.004

Li, X., Shi, W., Zhang, H., Peng, C., Wu, S., & Tong, W. (2025). The Agentic-AI Core: an AI-Empowered, Mission-Oriented core network for Next-Generation mobile telecommunications. *Engineering*. https://doi.org/10.1016/j.eng.2025.06.027

Lim, S., Schmälzle, R., & Bente, G. (2025). Artificial Social Influence via Human-Embodied AI Agent Interaction in Immersive Virtual Reality (VR): Effects of Similarity-Matching during health conversations. *Computers in Human Behavior Artificial Humans*, *5*, 100172. https://doi.org/10.1016/j.chbah.2025.100172

Martinez, D. R., & Kifle, B. M. (2024). *Artificial Intelligence: A Systems Approach from Architecture Principles to Deployment*. MIT Press eBooks, IEEE Xplore2. https://doi.org/10.7551/mitpress/14806.001.0001

Mohanarangan, S., Karthika, D., Moohambigai, B., & Sangeetha, R. (2024). Unleashing the Power of AI and Machine Learning: Integration Strategies for IoT Systems. *International Journal of Scientific Research in Computer Science and Engineering*, *12*(2), 25–32.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). https://doi.org/10.1145/259963.260288

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall. Google-Books-ID: CUVeMwAACAAJ.

Shamsudhin, N., & Jotterand, F. (2021). Social Robots and Dark Patterns: Where Does Persuasion End and Deception Begin? In F. Jotterand, & M. Ienca (Eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues* (pp. 89–110). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-74188-4_7

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. *arXiv:2210.05791*. https://doi.org/10.48550/arXiv.2210.05791

Small., C., & Lew, C. (2021). Mindfulness, moral reasoning and responsibility: Towards virtue in ethical decision-making. *Journal of Business Ethics*, *169*(1), 103–117. https://doi.org/10.1007/s10551-019-04272-y

Yang, J., Jimenez, C. E., Wettig, A., Lieret, K., Yao, S., Narasimhan, K., & Press, O. (2024). SWE-AGENT: Agent-Computer Interfaces Enable Automated Software Engineering. *arXiv:2405.15793*. https://doi.org/10.48550/arXiv.2405.15793

Yang, E., Garcia, T., Williams, H., Kumar, B., Ramé, M., Rivera, E., Ma, Y., Amar, J., Catalani, C., & Jia, Y. (2024). From barriers to tactics: A behavioural science-informed agentic workflow for personalized nutrition coaching. *arXiv:2410.14041*. https://doi.org/10.48550/arXiv.2410.14041

Author information

Gordon Bowen – DBA, Associate Professor, School of Management, Anglia Ruskin University
Address: East Road, CB1 1PT, Cambridge, United Kingdom
E-mail: gordon.bowen@aru.ac.uk
ORCID ID: https://orcid.org/0009-0007-4082-0336
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=56943078600
WoS Researcher ID: https://www.webofscience.com/wos/author/record/65121803
Google Scholar ID: https://scholar.google.com/citations?user=zm_Qgw4AAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – June 10, 2025
Date of approval – June 26, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:340.1:004.8
EDN: https://elibrary.ru/pltfwo
DOI: https://doi.org/10.21202/jdtl.2025.17

Агентный искусственный интеллект: правовые и этические вызовы автономных систем

Гордон Боуэн

Университет Англия Рёскин, Кембридж, Великобритания

Ключевые слова

автономность,
агентный искусственный интеллект,
искусственный интеллект,
ответственность,
право,
правовое регулирование,
программирование,
риск,
цифровые технологии,
этика

Аннотация

Цель: определить специфические правовые и этические проблемы агентного искусственного интеллекта и выработать рекомендации по созданию защитных механизмов для обеспечения ответственного функционирования автономных ИИ-систем.

Методы: исследование носит концептуальный характер и основано на системном анализе научной литературы по вопросам этики искусственного интеллекта, правового регулирования автономных систем и социального взаимодействия ИИ-агентов. В работе применяются сравнительный анализ различных типов ИИ-систем, исследование потенциальных рисков и преимуществ агентного искусственного интеллекта, а также междисциплинарный подход, интегрирующий достижения в сфере права, этики и компьютерных наук для формирования комплексного понимания проблематики.

Результаты: установлено, что агентный искусственный интеллект, обладая автономностью принятия решений и способностью к социальному взаимодействию, создает качественно новые правовые и этические вызовы по сравнению с традиционными ИИ-ассистентами. Выявлены основные категории потенциального вреда: прямое воздействие на пользователей через открытые и скрытые действия, манипулятивное влияние на поведение и кумулятивный вред от длительного взаимодействия. Определена необходимость распределения ответственности между тремя ключевыми субъектами: пользователем, разработчиком и владельцем системы агентного искусственного интеллекта.

Научная новизна: впервые проведен системный анализ этических аспектов агентного искусственного интеллекта как качественно нового класса автономных систем, отличающихся от традиционных ИИ-ассистентов степенью независимости и социальной интерактивности. Разработана типология потенциальных рисков социального взаимодействия с агентными интеллектуальными системами, и предложена концептуальная модель распределения правовой и этической ответственности в триаде «пользователь – разработчик – владелец».

Практическая значимость: результаты исследования формируют теоретическую основу для разработки этических принципов и правовых норм регулирования агентного искусственного интеллекта в условиях растущего рынка автономных интеллектуальных систем. Полученные выводы могут быть использованы законодателями при создании нормативной базы, разработчиками при проектировании защитных механизмов, а также организациями при внедрении агентных систем искусственного интеллекта в различных сферах экономической деятельности.

Для цитирования

Боуэн, Г. (2025). Агентный искусственный интеллект: правовые и этические вызовы автономных систем. *Journal of Digital Technologies and Law*, *3*(3), 431–445. https://doi.org/10.21202/jdtl.2025.17

Список литературы

Abuelsaad, T., Akkil, D., Dey, P., Jagmohan, A., & Vempaty, A. (2024). Agent-E: From Autonomous Web Navigation to Foundational Design Principles in Agentic Systems. *arXiv preprint arXiv:2407.13032*. https://doi.org/10.48550/arXiv.2407.13032

Acharya, D. B., Kuppan, K., & Ashwin, D. B. (2025). Agentic AI: Autonomous intelligence for complex goals – a comprehensive survey. In *IEEE Access (*vol. 13, pp. 18912–18936). https://doi.org/10.1109/ACCESS.2025.3532853

Alberts, L., Keeling, G., & McCroskery, A. (2024). Should agentic conversational AI change how we think about ethics? Characterising an interactional ethics centred on respect. *arXiv:2401.09082v2*. https://doi.org/10.48550/arXiv.2401.09082

Alberts, L., Lyngs, U., & Van Kleek, M. (2024). Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–25. https://doi.org/10.1145/3653693

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922

Beulen, E., Plugge, A., & van Hillegersberg. J. (2022). Formal and relational governance of artificial intelligence outsourcing. *Information System E Business Management*, *20*(4), 719–748. https://doi.org/10.1007/s10257-022-00562-7

Coghlan, S., Waycott, J., Lazar, A., & Neves, B. (2021). Dignity, Autonomy, and Style of Company: Dimensions Older Adults Consider for Robot Companions. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449178

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J., & Zitnik, M. (2024). Empowering biomedical discovery with AI agents. *Cell*, *187*(22), 6125–6151. https://doi.org/10.1016/j.cell.2024.09.022

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515.

Hasan, S. S., Fury, M. S., Woo, J. J., Kunze, K. N., & Ramkumar, P. N. (2025). Ethical Application of Generative Artificial Intelligence in Medicine. *Arthroscopy: Journal of Arthroscopic Related Surgery*, *41*(4), 874–885. https://doi.org/10.1016/j.arthro.2024.12.011

Hosseini, S., & Seilani, H. (2025). The Role of Agentic AI in Shaping a Smart Future: A Systematic review. *Array*, *26*, 100399. https://doi.org/10.1016/j.array.2025.100399

Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2024). AI Agents That Matter. *arXiv:2407.01502v1*.

Kocielnik, R., Langevin, R., George, J. S., Akenaga, S., Wang, A., Jones, D. P., Argyle, A., Fockele, C., Anderson, L., Hsieh, D. T., Kabir, Y., Duber, H., Hsieh, G., & Hartzler, A. L. (2021). Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain*. ACM, New York, NY, USA. https://doi.org/10.1145/3469595.3469599

Korzynski, P., Edwards, A., Gupta, M. C., Mazurek, G., & Wirtz, J. (2025). Humanoid robotics and agentic AI: reframing management theories and future research directions. *European Management Journa*l, *43*(4), 548–560. https://doi.org/10.1016/j.emj.2025.06.002

Kshetri, N. (2025). Transforming cybersecurity with agentic AI to combat emerging cyber threats. *Telecommunications Policy*, *49*(6), 102976. https://doi.org/10.1016/j.telpol.2025.102976

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a science of human-AI decision making: A survey of empirical studies. *arXiv:2112.11471*. https://doi.org/10.48550/arXiv.2112.11471

Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Network*, *145*, 80–89. https://doi.org/10.1016/j.neunet.2021.10.004

Li, X., Shi, W., Zhang, H., Peng, C., Wu, S., & Tong, W. (2025). The Agentic-AI Core: an AI-Empowered, Mission-Oriented core network for Next-Generation mobile telecommunications. *Engineering*. https://doi.org/10.1016/j.eng.2025.06.027

Lim, S., Schmälzle, R., & Bente, G. (2025). Artificial Social Influence via Human-Embodied AI Agent Interaction in Immersive Virtual Reality (VR): Effects of Similarity-Matching during health conversations. *Computers in Human Behavior Artificial Humans*, *5*, 100172. https://doi.org/10.1016/j.chbah.2025.100172

Martinez, D. R., & Kifle, B. M. (2024). *Artificial Intelligence: A Systems Approach from Architecture Principles to Deployment*. MIT Press eBooks, IEEE Xplore2. https://doi.org/10.7551/mitpress/14806.001.0001

Mohanarangan, S., Karthika, D., Moohambigai, B., & Sangeetha, R. (2024). Unleashing the Power of AI and Machine Learning: Integration Strategies for IoT Systems. *International Journal of Scientific Research in Computer Science and Engineering*, *12*(2), 25–32.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). https://doi.org/10.1145/259963.260288

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall. Google-Books-ID: CUVeMwAACAAJ.

Shamsudhin, N., & Jotterand, F. (2021). Social Robots and Dark Patterns: Where Does Persuasion End and Deception Begin? In F. Jotterand, & M. Ienca (Eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues* (pp. 89–110). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-74188-4_7

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. *arXiv:2210.05791*. https://doi.org/10.48550/arXiv.2210.05791

Small., C., & Lew, C. (2021). Mindfulness, moral reasoning and responsibility: Towards virtue in ethical decision-making. *Journal of Business Ethics*, *169*(1), 103–117. https://doi.org/10.1007/s10551-019-04272-y

Yang, J., Jimenez, C. E., Wettig, A., Lieret, K., Yao, S., Narasimhan, K., & Press, O. (2024). SWE-AGENT: Agent-Computer Interfaces Enable Automated Software Engineering. *arXiv:2405.15793*. https://doi.org/10.48550/arXiv.2405.15793

Yang, E., Garcia, T., Williams, H., Kumar, B., Ramé, M., Rivera, E., Ma, Y., Amar, J., Catalani, C., & Jia, Y. (2024). From barriers to tactics: A behavioural science-informed agentic workflow for personalized nutrition coaching. *arXiv:2410.14041*. https://doi.org/10.48550/arXiv.2410.14041

Сведения об авторе

Боуэн Гордон– доктор делового администрирования, доцент, школа менеджмента, Университет Англия Рёскин
Адрес: Великобритания, г. Кембридж, CB1 1PT, Ист Роуд
E-mail: gordon.bowen@aru.ac.uk
ORCID ID: https://orcid.org/0009-0007-4082-0336
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=56943078600
WoS Researcher ID: https://www.webofscience.com/wos/author/record/65121803
Google Scholar ID: https://scholar.google.com/citations?user=zm_Qgw4AAAAJ

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс
Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 10 июня 2025 г.
Дата одобрения после рецензирования – 26 июня 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.

Research article
UDC 34:004:346.6:004.8
EDN: https://elibrary.ru/ruzmxp
DOI: https://doi.org/10.21202/jdtl.2025.18

Legal Mechanisms for Distributing the Responsibility for the Harm Caused by Artificial Intelligence Systems

Dmitriy Aleksandrovich Kazantsev
Chamber of Commerce and Industry of the Russian Federation, Moscow, Russia

Keywords

artificial intelligence,
autonomy,
delictual dispositive capacity,
digital technologies,
law,
legal lliability,
legislation,
neural network,
risk-oriented approach,
robot

Abstract

Objective: to formulate proposals to form a system of subsidiary liability for harm resulting from the use of artificial intelligence systems.

Methods: the research is based on a comprehensive methodological basis, including the abstract logical method for theoretical understanding of the legal nature of artificial intelligence as an object of legal regulation; the method of comparison to analyze the Russian and European legislations on tort liability; generalization to systematize the existing concepts of responsibility distribution between subjects of law; and correlation analysis to identify the relationships between the typology of artificial intelligence systems and the mechanisms of legal responsibility for their functioning.

Results: the study summarizes and systematizes modern theoretical and legal concepts and regulations of the European Union and the Russian Federation on the distribution of subsidiary responsibility for the adverse effects of artificial intelligence. Potential subjects of responsibility were identified, as well as the key factors influencing the distribution of responsibility between them. A multidimensional matrix was developed for responsibility distribution between the subjects, taking into account their impact on the specific artificial intelligence system functioning and the systems typologization under the risk-based approach.

Scientific novelty: for the first time, an original concept is proposed, which combines the differentiation of the subjects' roles in terms of their real impact on the artificial intelligence results; the differentiation of artificial intelligence systems under the risk-based approach; and the system of legal

presumptions of responsibility distribution corresponding to the above two classifications. The novelty lies in the creation of a multidimensional matrix of subsidiary liability, which allows taking into account many factors when determining the subject of responsibility in each specific case of harm caused by artificial intelligence systems, which differs significantly from existing unilateral approaches to this issue.

Practical significance: the research conclusions and suggestions can be used to develop the doctrine of subsidiary responsibility in the field of artificial intelligence use, to develop and modify the legal norms regulating artificial intelligence. The proposed multidimensional matrix of responsibility distribution can serve as a theoretical basis for improving judicial practice in cases of compensation for damage caused by artificial intelligence systems, as well as for creating an effective balance between stimulating the development of AI technologies and ensuring the protection of the rights and legitimate interests of individuals and legal entities.

For citation

Kazantsev, D. A. (2025). Legal Mechanisms for Distributing the Responsibility for the Harm Caused by Artificial Intelligence Systems. *Journal of Digital Technologies and Law, 3*(3), 446–471. https://doi.org/10.21202/jdtl.2025.18

Contents

Introduction

The increasing use of artificial intelligence systems in everyday life, as well as in various sectors of the economy and even public administration, makes neural networks and other forms of the so-called weak artificial intelligence not just an experimental tool, but also a factor in legal relations. An important and one of the most socially significant facets of these relationships is delict relationships – in other words, obligations arising as a result of harm caused by the use of artificial intelligence. In a broader discourse, it is necessary to resolve the issue of assigning and distributing responsibility for the adverse consequences of AI use.

Robotization of industries and the increasing use of artificial intelligence in various aspects of everyday life is shifting the issue of the legal consequences of a robot causing

harm to humans from theory to practice. The lack of appropriate regulation creates a legal vacuum, which can potentially create a situation of lack of responsibility for a whole group of offenses.

This, in turn, will inevitably entail the desire of individuals and legal entities to avoid, as far as possible, involvement in such legal relationships in which a potential violation of their rights and legitimate interests may have no consequences. Simply put, the unresolved legal liability of AI is one of the key factors in the depopularization of the daily use of digital technologies, and therefore an important obstacle to their development.

Even today, this issue has ceased to be theoretical-legal. Unfortunately, the robotization of industries, jobs, and services demonstrates very real examples of how the ill-conceived use of artificial intelligence damages not only the rights and legitimate interests of individuals or legal entities, but also damages human health, and in some cases even leads to deaths. For example, robotization of vehicles and delivery services, medical diagnostics and personal data processing, while remaining an absolute convenience and a promising way of improving the living standards, has costs in the form of significant risks to the life and health of citizens.

At a higher level of generalization, one may consider AI systems to be threats to basic civil rights. "The obvious dangers include: infringements on privacy by covert surveillance (the European Court of Human Rights has already heard a number of cases on covert surveillance of employees in the workplace); dependence of the exercise of constitutional rights on the will of other actors (for example, providers); inadequate confidentiality when processing digitized personal information; additional costs for purchasing technical means and devices (for example, mandatory use of electronic diaries of schoolchildren in large families); linking to an electronic address to obtain official or banking information, etc." (Kovler, 2022).

Working in the digital environment in general and the consequences of AI actions in particular cannot and should not remain outside legal regulation. The inadmissibility of using artificial intelligence to intentionally cause harm to people and organizations, as well as preventing and minimizing the risks of negative consequences of using artificial intelligence are among the fundamental principles of AI development both in the Russian Federation and abroad[1].

As the very first approximation, practice leads us to choosing one of several simple solutions regarding the future legal regulation of AI.

The first and, it would seem, the most obvious solution is a complete ban on the use of any artificial intelligence systems as potentially dangerous to humans. Today, this danger is no longer purely speculative. It has been proven in practice. This means that

[1] Decree of the Russian President No. 490 of 10.10.2019 (edited as Decree of the Russian President of 124 of 15.02.2024). (2024). Garant. https://clck.ru/3NXX4N

it is necessary to eliminate any such systems from use in order to eliminate this threat to human life and health.

However, today artificial intelligence is not only a risk factor, but also a factor in improving quality of life and convenience of work for people. In the information age, a ban on an information processing tool would be as destructive as, for example, a ban on using cars or airplanes on the grounds that traffic accidents and plane crashes occur with deplorable regularity.

The ban itself is by no means an ordinary, but an extreme, exclusive instrument of legal regulation. Simply put, if it is possible to do without a ban in the field of law, then it is better to do without prohibitions. Any ban is nothing more than a stated crisis of public relations in one area or another and an imperfect social and legal regulation.

This does not mean that prohibitions are completely unnecessary and all of them are destructive. It is only important to use this instrument of legal regulation with great care and only if its absence can generate objectively greater risks than its presence.

A good example is the differentiation of risk factors for the AI use provided in a document known as the EU Artificial Intelligence Act[2]. According to this law, only those AI systems are subject to an unequivocal ban which are designed, for example, for biometric identification and categorization of people, building social rating systems, and the technologies that are directly aimed at destroying basic human and civil rights and freedoms. All other AI systems are subject only to more or less strict regulation, ranging from the maximum requirements for systems potentially capable of posing a threat to human life and health to the absence of any requirements for using AI in computer games.

Another extreme in solving the problem of legal regulation of the adverse effects of the AI use is to equate artificial intelligence technologies with force majeure circumstances or give them a similar status. At first glance, this is appropriate because the logic of information search and processing, even by weak AI, is not transparent to humans, which means that its solutions are far from predictable for humans.

However, the lack of full understanding does not mean that there is no possibility of influence. Continuing the analogy with a car, we can recall that in the 21st century, most motorists have rather vague understanding about the nuances of their car's design. Nevertheless, each of them is directly responsible for the consequences of managing the mechanism. In the same way, the impact on AI systems does not go beyond humans, even to the point of adjusting their algorithms in order to minimize the risk of erroneous decisions based on the results of big data processing.

In practice, any analogy between the work of an AI and a force majeure event will mean that there is no actual legal responsibility for the consequences of its work.

---

2 European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act). EUR-Lex. https://clck.ru/3NXX6n

Meanwhile, the exclusion of such responsibility obviously creates a space for abuse, including the use of artificial intelligence to commit crimes that would go unpunished in this case.

Since neither a complete ban nor complete impunity of AI seem reasonable and possible, the third approach pushes us to include AI in the circle of legal responsibility subjects, at least in cases where a person objectively did not and could not participate in an erroneous malicious decision made by artificial intelligence. This approach deserves closer consideration from a legal point of view.

1. Robot and human: bases of delictual dispositive capacity

Law, at least in its current form, is anthropocentric. It is a regulatory system created by humans for relations in human society. The development of law is a reflection of social development, and the evolution of law is subordinated to the evolution of both society and the position of an individual in society. Even when we are dealing with a legal fiction, assigning responsibilityto a legal entity, for example, practically means the occurrence of adverse consequences for specific individuals: managers, employees, owners, etc.

Apparently, legal regulation covers corporations, robots, other hardware and software systems, and mechanisms. Not limited to artifacts of human civilization, legal regulation in certain situations may even affect animals and plants (in particular, their breeding, turnover and handling). However, a non-human subject does not act as a subject of legal structures, not because of its “limitations” or “inferiority” in comparison with humans, but only because these structures emerged and were developing for many millennia precisely as a regulator of purely human behavior.

For example, the classical set of legal responsibility elements is commonly known to be a subject, a subjective side, an object and an objective side. While the presence of an object and an objective side is not eliminated by the fact of harm caused as a result of AI actions, the presence of the two remaining factors is debatable.

The European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics points to the increasing urgency of the liability for damage caused by artificial intelligence. At the same time, it notes that current legislation does not allow artificial intelligence to be liable even in cases where damage is caused to third parties[3]. Although the Resolution rather cautiously describes

[3] European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). (2018). Official Journal of the European Union, 252–257. https://clck.ru/3NXXAJ

the prospects for the legal subjectivity of artificial intelligence, its draft dated May 31, 2016, formulated several approaches to consolidate "the legal nature of artificial intelligence: to treat it as physicals persons, as legal entities, as animals or objects, or to create a new category with its own characteristics and consequences regarding the assignment of rights and obligations, including liability for damage"[4].

Hence, the issue was raised at the regulatory level that the use of AI could create a high or even unacceptable risk to people's lives, health, rights and legitimate interests, but this issue has not been resolved. Moreover, this risk is not determined by malicious AI actions. In this context, particularly relevant is the definition of "the legal nature of artificial intelligence: to treat it as physicals persons, as legal entities, as animals or objects, or to create a new category with its own characteristics and consequences regarding the assignment of rights and obligations, including liability for damage"[5].

Today, it seems premature to talk about assigning artificial intelligence the status of a legal entity. One has to agree that "the use of digital technologies using artificial intelligence at the current level of its development does not mean the emergence of new social relations that are qualitatively different from existing ones". A similar opinion is: "Artificial intelligence does not act as a digital legal entity in relations on the turnover of digital rights in the operator's information system. The operator uses digital technologies in entrepreneurship and applies elements of artificial intelligence in business models that do not generate digital legal relationships" (Andreev, 2021). In other words, artificial intelligence, being a tool for implementing traditional economic relations at a new technological level, does not generate fundamentally new legal relations so far.

This conclusion is also true from an ontological point of view. The skills of processing large amounts of information, including using self-learning technologies, do not create human-like thinking and consciousness. A fairly accurate and still relevant definition of AI is given in the above-mentioned Presidential Decree: "A set of technological solutions that allow simulating human cognitive functions (including self-learning and finding solutions without a predefined algorithm) and obtaining results comparable to at least the results of human intellectual activity when performing specific tasks"[6]. Today, the most realistic concept seems to be artificial intelligence as a software and hardware complex that has nothing in common with the human mind in terms of the essence of thinking, but is capable of solving generally similar or more complex tasks (Bokovnya et al., 2020).

---

4 Draft report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). (2016, May 31). Committee on Legal Affairs. https://clck.ru/3NXXCu

5 Nevejans, N. (2016). European Civil Law Rules in Robotics: Study. European Union. https://clck.ru/3NXXFs

6 Decree of the Russian President No. 490 of 10.10.2019 (edited as Decree of the Russian President of 15.02.2024 No. 124). (2024). Garant. https://clck.ru/3NXXHJ

Since similarity does not mean identity, assigning anthropomorphic features to a robot does not mean that it acquires identity with a human. Hence, given both the achieved technological and legal development, "obvious is the inconsistency of the proposal to recognize artificial intelligence as a legal personality similar to that of an individual, and despite using the human brain principles to build an artificial intelligence system, the principles of legal regulation of the human status cannot be applied to artificial intelligence" (Durneva, 2019).

Despite the rapid development of neural networks and robotics, the conclusion remains relevant that "giving robots (artificial intelligence systems) the legal entity status will not entail any explicit negative consequences in the foreseeable future. At the same time, the advantages of such a solution are not apparent compared to considering robots (artificial intelligence systems) as quasi-subjects of law. Based on Occam's philosophical principle not to multiply entities unless absolutely necessary, we believe that the introduction of such a fundamentally new legal entity as a robot (artificial intelligence system) into the legal sphere is premature (although it is possible that such a need may arise)" (Channov, 2022). Assigning human rights to AI would be just a purely incorrect extrapolation of human properties to AI (Duffy & Hopkins, 2013), without taking into account the specifics of either humans or artificial intelligence.

Putting this as simple as possible, one may argue that today it seems premature to assign AI the status of a legal entity. Yes, in many aspects, AI is qualitatively superior to humans both in the speed of information processing and in the volume of information being processed. However, assigning anthropomorphic features to a robot does not mean that it acquires identity with a human. Similarly, assigning AI the legal status of a human would only be an extrapolation of human properties to AI, which accounts the specifics of neither humans nor artificial intelligence.

The very phenomenon of delictual dispositive capacity is inextricably linked with the concept of legal personality. To simplify matters as much as possible, only a legal entity can be legally responsible. A party that is not a subject of law is not legally responsible. However, in a situation where the cause of tort obligations was the actions of an AI, i. e. an entity not being a subject of law, it is still necessary to have a mechanism of legal liability. It is obvious that even in the case of harm caused by artificial intelligence the subject of legal responsibility will be the subject of law or several such entities.

At the same time, it is important to emphasize that the legal principle of proportionality requires the imposition of responsibility on those persons whose actions directly or indirectly caused the occurrence of tort obligations.

2. Prevention of violations in the field of using AI

The theoretical and legal comprehension of the tort consequences of using artificial intelligence should not be limited to working with fait accompli and assigning responsibility for them. Such an approach appears to be inherently insufficient. It is not only possible, but

also necessary that the mechanisms for prevention, or rather, minimizing the risk of harm caused by artificial intelligence be included in the theoretical and legal justification, and then applied regulation.

To do this, first of all, it is necessary to theoretically substantiate, practically test (and, where possible and necessary, also normalize) the correct balance between the roles of humans and artificial intelligence in the implementation of business processes in economic relations. Legal regulation, in turn, should be adequate to these fundamental approaches. First of all, we should recognize that it is unacceptable to entrust AI with making key decisions on issues affecting the rights and legitimate interests of individuals and legal entities in any relationship. Such decisions must remain the exclusive prerogative of humans wherever possible.

The point here is not only and not so much in the above-mentioned principle of anthropocentrism of law. More precisely, this circumstance is due to the fact that a person is involved in the range of potential consequences of the robot's activity not as a subject, but as an object. For example, if a decision is made about an emergency shutdown of a power plant or an emergency discharge of water from a reservoir, even the most advanced algorithm, in the absence of complex additional settings, will evaluate the purely economic consequences of each possible solution. It may well turn out that the sudden flooding of a nearby village will be considered economically feasible by artificial intelligence in this situation. For other reasons, but with the same fatal logic, a decision may be made to hit a pedestrian or turn off life support systems. Postulating such principles as "a robot must protect a human" and "a robot cannot harm a human" at the level of basic AI algorithms cannot mitigate this risk: too often linear logic, divorced from ethics, will push to protect one person by harming another.

This brings us back to the concept of the subjective side of the offense. A highly qualified development team can create an algorithm that allows, to a certain extent, including human interests as the highest priority in the robot's decision-making mechanism. But even this will not endow the robot with the moral and ethical properties of a human being, imitate moral experiences and a psycho-emotional attitude to the action performed. Existing AI systems have such specific characteristics as non-transparent decision-making, autonomy, self-learning, and unpredictability in some cases from the viewpoint of human logic (Llorca, 2023).

One should remember: the methods of human thinking qualitatively differ from those of information processing by conditional artificial intelligence. These are two different types of information processing, each with its own advantages and disadvantages. Even if there are similar goals, it is inappropriate to transfer the properties of one type to another. The discussion about the possibility of artificial intelligence having consciousness as such remains beyond the scope of this article, but it is hardly possible to expect a uman-type consciousness from artificial intelligence.

Under these conditions, as a general rule, only a person, having proper expertise, developed ethical principles, and an anthropocentric axiology, can make decisions within the limits of their authority that affect the key rights and legitimate interests of another person. Therefore, today, artificial intelligence, as a general rule, can only be a source of data for an expert, and not a substitute for that expert. Moreover, the rule remains relevant: the less threats a potential AI error can pose to the life and health of individuals, the more appropriate it is to use AI.

If this rule is followed, fundamentally new legal structures are not required to regulate liability for harm resulting from the work of AI. The distribution of tort obligations may well be resolved using existing legal mechanisms. At the same time, the high autonomy of AI in decision-making, even if it does not place it among the subjects of law, requires thoughtful consideration of the AI specifics when adapting existing norms and principles of law. "Traditional approaches to legal responsibility attribution require significant adaptation, necessitating a more nuanced, multi-level framework that maintains clear accountability chains" (Tianran, 2024).

One of the most significant areas of adaptation is the regulation of responsibility in situations when it is impossible to preserve decision-making by humans, since AI autonomy is the very essence of automating this process.

A good example is a robotic taxi, an autopilot of a marine vessel, or a robot for microsurgical operations. The technological complexity of the tasks they perform directly determines the economic and practical expediency of removing a human from controlling these operations. Simply put, if there is a driver at the wheel, then a robot in such a taxi is no longer needed; and if a taxi is operated by a robot, then the presence of a driver negates the economic benefits of taxi robotization. It is for such situations that new legal mechanisms for regulating responsibility for the actions of artificial intelligence are relevant.

To a certain extent, in this situation, an analogy can be proposed between responsibility for the actions of an AI and those of an infant: both are not delictable but show high autonomy and limited predictability of their actions. A legal historian can even suggest an appeal to old, long-forgotten legal models. For example, the model of the relationship between paterfamilias and a slave in Roman law. Under this approach, "the legal status of AI becomes identical or close to that of Roman anthropomorphic collective organizations, or even more reduced – slaves, family members, children, including filius in potestate tua est" (Afanasyev, 2022). However, the simple principle "the master is legally responsible for all the actions of the slave" cannot be mechanically transferred to the model of distribution of responsibility for the consequences of the robot's actions.

This does not mean that modern artificial intelligence is more complex in psycho-emotional terms than, for example, an ancient Roman gladiator. On the contrary, when

investigating the causes of a decision by artificial intelligence, we can more easily identify at least several key actors, and each of these subjects will be an individual, a legal entity, or a group of persons. "Relations using artificial intelligence are always relations between subjects of law or about objects of law. In any case, these are relationships that are initiated and programmed at one stage or another by a person – a subject of law with varying degrees of responsibility (including within the activities of legal entities). A person's will to certain actions of artificial intelligence can be expressed in varying degrees: from the AI actions under the full control of a human will, to autonomous AI actions allowed and realized within their possible limits and consequences by a person (group of persons)" (Shakhnazarov, 2022).

This possibility, in turn, can and should become the basis for building a model of subsidiary responsibility (Laptev, 2019) for the actions of artificial intelligence between the subjects of law that could directly or indirectly influence such actions.

3. Multidimensional matrix of subsidiary responsibility for AI functioning

The issue of tort obligations arising from the activities of AI has been raised in theoretical-legal works for several years now (Bertolini, 2013). Special studies, as a rule, contain conclusions about subsidiary liability or a liability matrix; they serve as the basis on which the question of imposing adverse legal consequences is decided individually in each case, taking into account a set of facts (Bokovnya et al., 2020). This implies subjects such as the AI owner, user, developer, and third parties. It seems appropriate to base the matrix of distribution of legal responsibility for the consequences of AI actions on the balance of the rights, duties and responsibilities of these subjects.

Apparently, we cannot apply criminal or administrative penalties to AI. Any responsibility measure applied to AI will in any case entail adverse consequences for its user: for example, an administrative ban on AI operation for a certain period will mean costs not for the AI, but only for the entity that used this AI system in its business activities. "When considering AI liability, it is relevant to talk primarily about tort liability; that is, liability measures should be established as a reaction to the harm that AI can or does cause. At the same time, this is not always about linear responsibility, i.e. the responsibility of a person for the harm that they caused, but rather about combined responsibility, i.e. when, in addition to the harm-doer, other actors may be called to account" (Philipp, 2023).

A theoretical and legal solution to the issue of allocating responsibility for AI actions that have caused damage to individuals and/or legal entities is necessary as a basis for regulating the practical aspects of the consequences of such liability. "A fundamental issue underlying AI liability is responsibility fragmentation. Unlike traditional tools that function under direct human control, AI-driven systems operate autonomously based

on algorithmic decision-making. In product liability cases, manufacturers are generally held accountable for design flaws, but what happens when an AI system "learns" harmful behavior over time? Some legal scholars advocate for strict liability on manufacturers, similar to pharmaceutical industry regulations, while others propose shared responsibility models that include software developers, operators, and even end-users"[7].

To resolve this issue, it is necessary, first of all, to identify a range of legal entities that are delictable and able to actually compensate for the damage caused by AI errors. For example, according to the long-held opinion of R. Leenes and F. Lucivero, the responsibility for the harm caused by AI lies with the person who programmed it or the person responsible for its operation, within the limits established by law (Leenes & Lucivero, 2014). At the same time, the principle of legal proportionality requires the existence of a causal relationship between the action (inaction) of such persons and the occurrence of the said harm. From this point of view, the following groups of potential subjects of responsibility can be distinguished:

1. AI developer.
2. AI owner.
3. AI user.
4. Third parties.

This list identifies the groups in which subgroups can be distinguished. For example, in practice, it is possible to separate AI customers from its developers; among third parties, one may distinguish those who had a direct impact on AI algorithms from those who posted false information in the public domain, which caused erroneous AI decisions, etc. In any case, by supplementing the generalized list above with two forms of guilt, one gets a two-dimensional matrix of subsidiary responsibility for the consequences of AI decisions (Table 1).

Table 1. Basic matrix of culpable responsibility for AI functioning

Subject of responsibility	Intent	Negligence
Developer		
Owner		
User		
Third parties		
Regulatory bodies		

Subsidiary liability is formed as a result of clarifying the presence and nature of guilt in each case. The matrix allows taking into account many factors: for example, whether the user followed the instructions of the developers during the AI operation; whether there were any restrictions in this particular AI model; whether they were brought to the

[7] Upadhyay, Sh. (2025, March 6). Navigating Liability in Autonomous Robots: Legal and Ethical Challenges in Manufacturing and Military Applications. https://clck.ru/3NXXNJ

attention of the user; whether the AI system underwent proper training (and if not, whether the damage was caused by the fault of the developer or of the agent who provided data for training); to what extent the owner could control the AI operation or the user's actions, etc.

For example, based on the results of an expert examination (and, if necessary, investigative actions), it can be proved that adverse consequences have arisen due to fundamental design flaws. "When an artificial intelligence system is purchased embedded in other goods (for example, in a car), it seems unlikely that such contractual exclusions (for example, between the car manufacturer and the provider of artificial intelligence software) can be successfully transferred to the car buyer. At the same time, of interest is the idea of establishing the boundaries of developers' responsibility for defects in the creation of released artificial intelligence systems" (Kharitonova et al., 2022)[8].

The thesis on the responsibility limits brings us to the question of legal presumptions in the field of regulating responsibility for AI actions. For example, one can distribute such presumptions of responsibility among the above-listed entities: "Primary responsibility rests with deploying organizations and system operators who maintain direct control over implementation, requiring them to ensure proper function, monitor performance, and implement necessary safeguards while maintaining comprehensive documentation. Secondary responsibility extends to system developers and manufacturers, encompassing technical standards compliance, safety features, and documentation requirements, including transparent decision-making processes and clear audit trails. Tertiary responsibility belongs to oversight bodies and regulatory authorities, who must establish standards, conduct regular audits, and maintain effective enforcement mechanisms. This layered framework ensures comprehensive coverage of responsibility while maintaining clear accountability chains throughout the system's lifecycle" (Tianran, 2024).

By supplementing our two-dimensional matrix with such presumptions, we obtain the following logic for allocating responsibility for AI actions (Table 2).

Table 2. Variant of the parity of responsibility for AI functioning

Order of bringing to responsibility	Guilt	
	Intent	Negligence
1. Owner		
2. Developer		
3. Customer		
4. User		
5. Regulatory and controlling bodies		
6. Information provider		
7. Third parties		

[8] Naumov, V. B., Chekhovskaya, S. A., Braginets, A. Yu., & Mayorov, A. V. (2021). Legal aspects of using artificial intelligence: current problems and possible solutions: report of the Higher School of Economics. Moscow.

At the same time, "a high degree of AI autonomy cannot serve as a basis for reducing the responsibility of developers and manufacturers. If an AI developer has a greater degree of control over the functioning of an AI system than the system manufacturer, owner, or user, this should increase the developer's responsibility for causing harm. This principle can be presented in a more universal interpretation: the degree of control over the AI system functioning is proportional to the responsibility for causing harm"[9].

Without using presumptions, the issue of allocating responsibility for AI actions will indeed be difficult to resolve in many cases. However, such a linear distribution of presumptions seems to be a simplification. It is obvious to a practicing lawyer that this matrix cannot cover all possible combinations of responsibilities that potentially arise when using AI. Guiltless compensation for damages (both contractual and non-contractual) remains beyond it, as well as the design of the source of increased danger.

Yes, as a general rule of Part 2 of Article 1064 of the Russian Civil Code, a person who has caused harm is exempt from compensation if they prove that the harm was caused no through their fault. However, there are exceptions to this rule. One of them is established by Art. 1079 of the Civil Code: persons who own a source of increased danger are obliged to compensate for the damage caused, regardless of the presence or absence of guilt, if the damage was caused by this particular source of increased danger. The Supreme Court of the Russian Federation explains that a source of increased danger is "any activity that creates an increased likelihood of harm due to the inability of a person to fully control it"[10].

Intuitively, AI falls under this definition, because it is autonomous in its decisions, is not fully controlled by humans, and is capable of harming individuals and legal entities. "The source of increased danger can be recognized through the following criteria: 'activity', 'action', and 'harmfulness'. Having identified the necessary criteria, one has to find out whether the AI meets the specified requirements. The categories of 'activities' and 'actions' that pose a risk of harm are confirmed by the technically complex structure of the AI technology, as well as by the autonomous choice of a strategy for completing the task. The criterion of 'harmfulness' is revealed through the areas in which AI technology can be used. For example, the use of artificial intelligence in medicine in determining the diagnosis or in unmanned vehicle control presupposes the possibility of harming surrounding subjects. Thus, one may conclude that artificial intelligence can be recognized as a source of increased danger"[11].

---

9 Naumov, V. B., Chekhovskaya, S. A., Braginets, A. Yu., & Mayorov, A. V. (2021). Legal aspects of using artificial intelligence: current problems and possible solutions: report of the Higher School of Economics. Moscow..

10 On the application by courts of civil legislation, regulating relations on obligations due to harm to the life or health of a citizen: Resolution of the Plenum of the Supreme Court of the Russian Federation No. 1 of 26.01.2010, cl. 18.

11 Pozdnyakova, M. (2025, April 3). Recognition of artificial intelligence as a source of increased danger: realities and prospects. Delovoy profil. https://clck.ru/3NXXbo

Indeed, "many AI systems (unmanned vehicles, drones, surgical robots, etc.) can be classified as sources of increased danger, and their use as activities that pose an increased risk to others" (Izhaev & Kuteynikov, 2024). This thesis is the starting point for concretizing the correct but abstract thesis about the individual decision on the subsidiary responsibility allocation in each specific case, taking into account a set of factors. "The incorporation of new AI systems into law will require considering their recognition as a source of increased danger" (Antonov, 2020); in this context, the owner of artificial intelligence appears to be the owner of a source of increased danger. By default, the owner is responsible for the adverse effects of artificial intelligence activities, but only until the guilt of other persons is proven.

Summarizing the above theses and partially complementing them, we may state that "according to the criterion of applying the existing legal regulation to harm caused by AI systems, the following approaches are possible:

– lability for harm caused by a source of increased danger;

– liability for damage caused due to defects (defects) of the product;

– guiltless liability for harm caused by extremely dangerous activities;

– analogy to the norms on liability for harm caused by animals. In particular, there are some similarities between robots and animals. For example, both robots and animals can act independently of their owners, perceive the environment and perform actions depending on it;

– analogy to the norms on liability for harm caused by employees. The employer's liability for harm caused by an employee to third parties is related to the actions of the employee that caused harm, which they committed as part of work duties;

– analogy to the norms on liability for harm caused by children"[12].

All of these approaches are correct in their own way, but each of them can only be fully applied to individual AI use cases. After all, the concept of a source of increased danger today is no longer able to fully cover the practice of using AI. "Obviously, there are different types of AI systems, from robot vacuum cleaners to autonomous drones used in weapons. A large number of primitive AI systems will not have characteristics capable of causing any significant harm to humans. In this regard, the default usage of Art. 1079 of the Russian Civil Code and equating all AI systems to sources of increased danger is controversial. In part, we can agree with the expediency of detailing criteria for sources of increased danger in relation to AI systems. It should be borne in mind that in practice, the so-called activity approach can be used to determine the source of increased danger in

[12] Naumov, V. B., Chekhovskaya, S. A., Braginets, A. Yu., Mayorov, A. V. (2021). Legal aspects of using artificial intelligence: current problems and possible solutions: report of the Higher School of Economics. Moscow.

a particular situation. Its essence is reflected in the resolution of the Plenum of the Russian Supreme Court No. 138 dated 26.01.2010. The document states that within the meaning of Art. 1079 of the Russian Civil Code, as a source of increased danger is any activity that creates an increased likelihood of harm due to the inability of a person to fully control it, as well as activities related to the use, transportation, storage of objects, substances, and other industrial, economic, or other facilities with the same properties. This interpretation allows the court to determine on a case-by-case basis whether an AI system is a source of increased danger" (Izhaev & Kuteynikov, 2024).

Hence, not every AI system is a source of increased danger. This means that the application of the presumptions briefly listed above is subject-object in nature. In other words, it depends not only on the status of the legal responsibility subject, but also on AI system as the object, whose work results imply responsibility.

Here we return to the risk-based classification of the EU AI Act, mentioned at the very beginning of this article. To implement a risk-based approach, the Act identifies four groups of AI systems, which vary depending on the purpose:

1. Unacceptable risk category: biometric identification and categorization of people, social rating system, etc.;

2. High risk category: the use of AI which may create a direct threat to human life and health, for example, in transport or medicine;

3. Low risk category: the use of chatbots, the use of neural networks to create information content;

4. Minimum risk category: video games, assistants, recommendation systems.

The classification certainly deserves further development. For example, within the group of high-risk systems, it is worth separating systems with a risk to life and health, on the one hand, and systems with a risk to the property interests of individuals and legal entities, on the other. General purpose and specialized systems can be distinguished within low-risk systems. In any case, when allocating subsidiary responsibility, it is necessary to take into account not only the impact that tort-related subjects of law have on AI functioning, but also the nature of the specific AI system.

For example, "in cases where harm is caused by a high-risk AI system, it is advisable to use strict developer responsibility. This is due to the fact that such AI systems, by their very nature, can have a significant negative impact on human rights and freedoms, and therefore the latter should have increased protection guarantees" (Izhaev & Kuteynikov, 2024). The implementation of this approach shifts the discourse from the concept of a source of increased danger towards establishing its own system of presumptions for each group of AI systems, distinguished according to the risk-oriented principle.

As a result, we have a multidimensional matrix that takes into account at least the following parameters:

1. The role of a delictable legal entity in AI functioning;
2. The form of guilt and the existence of grounds for guiltless responsibility;
3. The category of the AI system from the viewpoint of the risk-oriented approach.

In such a multidimensional matrix, we are no longer limited to categorical statements that "the owner is primarily responsible for AI" or "the developer is responsible for AI, and we hold everyone else accountable only after the absence of the developer's guilt is proven".

Yes, such a multidimensional matrix is complicated. However, it is the only opportunity to achieve the balance of rights and duties and to maintain the balance of economic interests. After all, while the complete absence of legal regulation of liability can impede the AI industry development, excessive under-elaborated regulation is quite capable of creating similar problems. Simply put, if we presume the responsibility of the owner for the AI actions, then no one will want to buy such systems. If developers are held accountable by default, then few people will want to develop them.

One should agree that "in the light of the possible imposition of responsibility on developers, it is necessary, at least at the initial stages, to provide a balanced system of immunities for them, adding mandatory liability insurance, as well as registration of AI systems. If AI is recognized as a subject, it is possible to establish a regime of combined responsibility, when both the AI creator and owner or another subject can bear subsidiary responsibility"[13]. It seems that such subsidiary responsibility can be implemented most effectively within the framework of the multidimensional matrix proposed above.

Despite its complexity, the multidimensional matrix will allow not only to include all cases of AI use in the legal regulation, but also to take into account the variability, changeability and specific combinations of various algorithms. For example, an AI owner becomes the first candidate for subsidiary liability in the event of harm due to the use of "general-purpose AI systems, which are characterized by the ability to solve a wide range of tasks. As a general rule, they should be defined as low-risk AI systems. However, if they are used in high-risk products as a result of 'fine-tuning', then such systems should also be recognized as high-risk with corresponding consequences in resolving disputes arising from harm" (Philipp, 2023). As for high-risk systems, the presumption of responsibility can be placed primarily on customers and developers.

Conclusions

Law as a phenomenon and legal institutions as its manifestations do not develop in a vacuum of purely theoretical constructions, but only in the developing practice of economic relations. In this sense, the theoretical understanding of technical and economic realities follows the

[13] Naumov, V. B., Chekhovskaya, S. A., Braginets, A. Yu., Mayorov, A. V. (2021). Legal aspects of using artificial intelligence: current problems and possible solutions: report of the Higher School of Economics. Moscow.

emergence of these realities. However, without theoretical understanding, neither systemic cognition nor professional regulation of new relationships are possible.

Today, due to the level of technology development and the involvement of innovative technologies in economic relations, the issue of AI responsibility is no longer only theoretical-legal but is rather of practical importance. Robots today can not only benefit, but also cause harm to both individuals and legal entities. Moreover, "the use of algorithmic systems poses particular threats to personal and political rights – the right to privacy, freedom of expression, and the right to participate in state governance through democratic procedures. In addition, due to the fact that algorithms and artificial intelligence technologies based on them process information from the external environment, the rights of personal data subjects should be under special protection in an algorithmic society" (Pibaev & Simonova, 2020).

The lack of special regulation creates a legal vacuum, which potentially means that there is no responsibility for a group of offenses. This, in turn, is a key factor in the depopularization of using digital technologies, and therefore an important obstacle to their development. However, ill-conceived regulation can become the same, if not a more significant obstacle to using AI in everyday and industrial matters.

The first but important step towards practical regulation should be the theoretical-legal elaboration of issues of responsibility for the consequences of AI use. "There is still no clear understanding of how to resolve the problems of imposing non-contractual civil liability for harm caused by AI systems. On the one hand, regulation should stimulate the AI sector development and not contain excessively burdensome provisions for developers and professional operators. On the other hand, it is necessary to ensure a high level of protection of the rights of humans and society, since the latter will obviously be the weak side in such disputes. Thus, it is obvious that searching for optimal and adequate approaches to the legal regulation of legal liability is urgent" (Izhaev & Kuteynikov, 2024).

Today, the issue of responsibility for the consequences of AI actions can be resolved positively, since the intentional or at least careless fault of "artificial intelligence intermediaries (developers and users) in the event of harm by an artificial intelligence system can be quite probable, legally and expertly provable" (Ivliev & Egorova, 2022). This means that the principle of "delineating the responsibilities of organizations that develop and use artificial intelligence technologies based on the nature and degree of harm caused" already seems feasible[14].

[14] Decree of the Russian President No. 490 of 10.10.2019 (edited as Decree of the Russian President of 15.02.2024 No. 124). (2024). Garant. https://clck.ru/3NXXig

However, the possibility of a positive solution to the issue does not mean that it is easy to solve. First of all, it is necessary to rely on the following fundamental assumptions:

1. The current levels of development of both law and artificial intelligence technologies do not allow considering a robot as a subject of legal relations or of legal responsibility.

2. The impossibility of recognizing the delictability of artificial intelligence does not mean that it must be recognized as force majeure or that it cannot be held responsible for the consequences of artificial intelligence actions.

3. Responsibility for the consequences of AI actions is distributed between its creators, owners, users and other persons involved in using the robot, in the extent that they affect the results of the artificial intelligence functioning.

4. The combination of liability, including subsidiary liability, in each specific case depends both on the type of the AI system and on whose actions influenced the AI's decision which resulted in the tort obligations.

For example, "the user or owner may be held liable if the instructions for using artificial intelligence are violated, especially in situations where the user was informed of specific requirements for the system operation. If we are talking about the user or the owner, then the model of responsibility for harm caused by a source of increased danger is the closest to this type of relationship. The data provider is responsible if the damage occurred when the system was still being trained, or if low-quality data was provided. It should also be borne in mind that the artificial intelligence system can be released with an open source code. In this case, experts speak of holding programmers accountable. Also, in some cases, if the damage is caused by deep-seated problems of the artificial intelligence system, the question arises of holding the designer or manufacturer of the artificial intelligence system accountable. Since artificial intelligence systems often operate in the aforementioned 'black box' paradigm, in some cases it may be impossible to identify the person by whose will or negligence the harm was caused" (Shakhnazarov, 2022)[15].

In these circumstances, a positive solution to the issue of subsidiary liability is impossible without applying the legal structures that are close (but not necessarily identical) to the concept of a source of increased danger. At the same time, it seems that such a structure, applicable in the field of responsibility for AI decisions, should not be limited only to the guiltless responsibility.

Somewhat simplifying, we can propose a system of responsibility presumption. Within this system, the guilt of each of the delictable legal entities is investigated in a top-down manner. Only in the case when it is objectively impossible to establish guilt, guiltless responsibility is applied.

[15] Naumov, V. B., Chekhovskaya, S. A., Braginets, A. Yu., & Mayorov, A. V. (2021). Legal aspects of using artificial intelligence: current problems and possible solutions: report of the Higher School of Economics. Moscow.

At the same time, the hierarchy of presumptions depends on the AI category within a risk-based approach. According to it, at least the following categories of AI can be distinguished:

1. High-risk AI that can pose a threat to human life and health;
2. High-risk AI that can pose a threat to the property of individuals and legal entities;
3. High-risk AI that can create a threat of disclosure of personal data and other information with limited access;
4. Medium-risk AI that can pose a threat to the proper conduct of business operations;
5. Medium-risk AI that can pose a threat to production processes and the functioning of infrastructure facilities;
6. Medium-risk AI of general purpose;
7. Low-risk AI.

For each of these categories, an individual system of responsibility presumptions is built for the following subjects:

1. AI owner;
2. AI customer;
3. AI developer;
4. AI user;
5. Regulatory and controlling bodies;
6. Providers of information for AI;
7. Third parties.

The proposed multidimensional matrix of responsibility for harm caused by AI actions is schematically presented in Table 3, where, for each cell in the indicated order, the question of intentional or negligent guilt is first investigated and then the possibility of guiltless liability.

Table 3. Multidimensional matrix of responsibility for AI functioning

Categories of AI systems	Subjects of responsibility						
	Owner	Customer	Developer	User	Regulatory bodies	Information provider	Third parties
High-risk AI that can pose a threat to human life and health	3	1	2	6	4	5	7
High-risk AI that can pose a threat to the property of individuals and legal entities	3	1	2	4	5	6	7
High-risk AI that can create a threat of disclosure of personal data and other information with limited access	4	1	2	3	6	5	7
Medium-risk AI that can pose a threat to the proper conduct of business operations	1	4	5	2	3	6	7
Medium-risk AI that can pose a threat to production processes and the functioning of infrastructure facilities	1	3	4	2	6	5	7
Medium-risk AI of general purpose	1	4	3	2	5	7	6
Low-risk AI	4	3	2	1	5	7	6

Thus, the formal inability to impose a punishment or other measure of legal responsibility on a robot today does not at all prevent the full inclusion of relations using artificial intelligence technologies in the sphere of legal regulation, including in terms of the legal consequences of harm. These innovative technologies require significant development of legal regulation, but they do not create either new legal institutions or fundamentally new legal structures. This means that with a proper approach to the essential understanding of the technological component, such regulation can be successfully implemented within the existing legal system.

References

Afanasyev, S. F. (2022). On the problem of substantive and procedural legal personality of artificial intelligence. *Vestnik Grazhdanskogo Protsessa*, *3*, 12–31. https://doi.org/10.24031/2226-0781-2022-12-3-12-31

Andreev, V. K. (2021). Acquiring and exercising rights of a legal entity with the use of artificial intelligence. *Predprinimatelskoe Pravo*, *4*, 11–17. https://doi.org/10.18572/1999-4788-2021-4-11-17

Antonov, A. A. (2020). Artificial intelligence as a source of increased danger. *Yurist*, *7*, 69–74. https://doi.org/10.18572/1812-3929-2020-7-69-74

Bertolini, A. (2013). Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules. *Law, Innovation and Technology*, *5*. https://doi.org/10.5235/17579961.5.2.214

Bokovnya, A. Y. et al. (2020). Legal Approaches to Artificial Intelligence Concept and Essence Definition. *Revista San Gregorio*, *41*, 115–121. https://doi.org/10.36097/rsan.v1i41.1489

Channov, S. E. (2022). Robot (artificial intelligence system) as a subject (quasi-subject) of law. *Actual Problems of Russian Law*, *12*, 94–109. https://doi.org/10.17803/1994-1471.2022.145.12.094-109

Duffy, S. H., & Hopkins, J. P. (2013). Sit, Stay, Drive: The Future of Autonomous Car Liability. *SMU Science & Technology Law Review*, *16*.

Durneva, P. N. (2019). Artificial intelligence: an analysis from the standpoint of the classical legal capacity theory. *Grazhdanskoe Pravo*, *5*, 30–35. https://doi.org/10.18572/2070-2140-2019-5-30-33

Ivliev, G. P., & Egorova, M. A. (2022). Legal issues of the legal status of artificial intelligence and products created by artificial intelligence systems. *Zhurnal Rossiyskogo Prava*, *6*, 32–46. https://doi.org/10.12737/jrl.2022.060

Izhaev, O. A., & Kuteynikov, D. L. (2024). Artificial intelligence systems and non-contractual civil liability: a risk-based approach. *Lex russica*, *77*(6), 23–34. https://doi.org/10.17803/1729-5920.2024.211.6.023-034

Kharitonova, Yu. S., Savina, V. S., & Pagnini, F. (2022). Civil liability in the development and application of artificial intelligence and robotic systems: basic approaches. *Vestnik Permskogo Universiteta. Yuridicheskie Nauki*, *58*, 683–708. https://doi.org/10.17072/1995-4190-2021-58-683-708

Kovler, A. I. (2022). Anthropology of human rights in the digital age (experience of comparative legal method). *Zhurnal Rossiyskogo Prava*, *12*, 5–29. https://doi.org/10.12737/jrl.2022.125

Laptev, V. A. (2019). Artificial intelligence and liability for its work. Law. *Journal of the Higher School of Economics*, *2*, 79–102. https://doi.org/10.17-323/2072-8166.2019.2.79.102

Leenes, R., & Lucivero, F. (2014). Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design. *Law, Innovation and Technology*, *6*(2), 194–222.

Llorca, D. F. (2023). Liability Regimes in the Age of AI: a Use-Case Driven Analysis of the Burden of Proof. *Journal of Artificial Intelligence Research*, *76*, 613–644. https://doi.org/10.48550/arXiv.2211.01817

Philipp, H. (2023). The European AI liability directives — Critique of a half-hearted approach and lessons for the future. *Computer Law & Security Review*, *51*, 1–42. https://doi.org/10.1016/j.clsr.2023.105871

Pibaev, I. A., & Simonova, S. V. (2020). Algorithms in the mechanism of implementation of constitutional rights and freedoms: challenges in the digital age. *Sravnitelnoye Konstitutsionnoye Obozreniye*, *6*, 31–50. https://doi.org/10.21128/1812-7126-2020-6-31-50

Shakhnazarov, B. A. (2022). Legal regulation of relations using artificial intelligence. *Actualniye Problemy Rossiyskogo Prava*, *9*, 63–72. https://doi.org/10.17803/1994-1471.2022.142.9.063-072

Tianran, L. (2024). Research on Legal Responsibility Attribution for Autonomous Systems: An AI Governance Perspective. *Science of Law Journal*, *3*(7), 166–174. https://doi.org/10.23977/law.2024.030722

Author information

Dmitriy A. Kazantsev – Cand. Sci. (Law), member of the Council for developing purchases, Chamber of Commerce and Industry of the Russian Federation, Moscow, Russia
Address: 6/1c1 Ilyinka Str., 109012, Moscow, Russia
E-mail: info@dkazantsev.ru
ORCID ID: https://orcid.org/0000-0003-2182-5776
RSCI Author ID: https://elibrary.ru/author_items.asp?authorid=1149755

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – May 22, 2025
Date of approval – June 4, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:346.6:004.8
EDN: https://elibrary.ru/ruzmxp
DOI: https://doi.org/10.21202/jdtl.2025.18

Правовые механизмы распределения ответственности за вред, причиненный системами искусственного интеллекта

Дмитрий Александрович Казанцев

Торгово-промышленная палата Российской Федерации, Москва, Россия

Ключевые слова

автономность,
деликтоспособность,
законодательство,
искусственный интеллект,
нейронная сеть,
право,
рискориентированный подход,
робот,
цифровые технологии,
юридическая ответственность

Аннотация

Цель: формулировка предложений по формированию системы субсидиарной ответственности субъектов за вред, ставший результатом использования систем искусственного интеллекта.

Методы: исследование базируется на комплексной методологической основе, включающей применение абстрактно-логического метода для теоретического осмысления правовой природы искусственного интеллекта как объекта правового регулирования, метода сравнения для анализа подходов российского и европейского законодательства к регулированию деликтной ответственности, методов обобщения для систематизации существующих концепций распределения ответственности между субъектами права, а также корреляционного анализа для выявления взаимосвязей между типологией систем искусственного интеллекта и механизмами правовой ответственности за их функционирование.

Результаты: в ходе исследования обобщены и систематизированы современные теоретико-правовые представления и нормативные акты Европейского союза и Российской Федерации о вариантах распределения субсидиарной ответственности за неблагоприятные последствия работы искусственного интеллекта. Определены потенциальные субъекты ответственности и выявлены ключевые факторы, влияющие на распределение ответственности между ними. Разработана многомерная матрица распределения ответственности между субъектами, учитывающая влияние каждого из них на работу конкретной системы искусственного интеллекта и типологизацию самих систем с точки зрения рискориентированного подхода.

Научная новизна: в работе впервые предложена авторская концепция, сочетающая дифференциацию ролей субъектов с точки зрения их реального влияния на результаты работы искусственного интеллекта, дифференциацию самих систем искусственного интеллекта согласно

рискориентированному подходу и соответствующую двум указанным классификациям систему правовых презумпций распределения ответственности. Новизна заключается в создании многомерной матрицы субсидиарной ответственности, которая позволяет учитывать множество факторов при определении субъекта ответственности в каждом конкретном случае причинения вреда системами искусственного интеллекта, что существенно отличается от существующих односторонних подходов к данной проблематике.

Практическая значимость: выводы и предложения исследования могут быть использованы для развития доктрины субсидиарной ответственности в области использования искусственного интеллекта, разработки и модификации норм права, посвященных регулированию искусственного интеллекта. Предложенная многомерная матрица распределения ответственности может служить теоретическим основанием для совершенствования судебной практики по делам о возмещении вреда, причиненного системами искусственного интеллекта, а также для создания эффективного баланса между стимулированием развития ИИ-технологий и обеспечением защиты прав и законных интересов физических и юридических лиц.

Для цитирования

Казанцев, Д. А. (2025). Правовые механизмы распределения ответственности за вред, причиненный системами искусственного интеллекта. *Journal of Digital Technologies and Law*, *3*(3), 446–471. https://doi.org/10.21202/jdtl.2025.18

Список литературы

Андреев, В. К. (2021). Приобретение и осуществление прав юридического лица с использованием искусственного интеллекта. *Предпринимательское право*, *4*, 11–17. EDN: https://elibrary.ru/nnesjs. DOI: https://doi.org/10.18572/1999-4788-2021-4-11-17

Антонов, А. А. (2020). Искусственный интеллект как источник повышенной опасности. *Юрист*, *7*, 69–74. EDN: https://elibrary.ru/dwhttx. DOI: https://doi.org/10.18572/1812-3929-2020-7-69-74

Афанасьев, С. Ф. (2022). К проблеме материальной и процессуальной правосубъектности искусственного интеллекта. *Вестник гражданского процесса*, *3*, 12–31. EDN: https://elibrary.ru/fjaogm. DOI: https://doi.org/10.24031/2226-0781-2022-12-3-12-31

Дурнева, П. Н. (2019). Искусственный интеллект: анализ с точки зрения классической теории правосубъектности. *Гражданское право*, *5*, 30–35. EDN: https://elibrary.ru/nckitd. DOI: https://doi.org/10.18572/2070-2140-2019-5-30-33

Ивлиев, Г. П., Егорова, М. А. (2022). Юридическая проблематика правового статуса искусственного интеллекта и продуктов, созданных системами искусственного интеллекта. *Журнал российского права*, *6*, 32–46. EDN: https://elibrary.ru/anagtu. DOI: https://doi.org/10.12737/jrl.2022.060

Ижаев, О. А., Кутейников, Д. Л. (2024). Системы искусственного интеллекта и внедоговорная гражданско-правовая ответственность: риск-ориентированный подход. *Lex russica*, *77*(6), 23–34. EDN: https://elibrary.ru/xvabso. DOI: https://doi.org/10.17803/1729-5920.2024.211.6.023-034

Ковлер, А. И. (2022). Антропология прав человека в цифровую эпоху (опыт сравнительного анализа). *Журнал российского права*, *12*, 5–29. EDN: https://elibrary.ru/drklnh. DOI: https://doi.org/10.12737/jrl.2022.125

Лаптев, В. А. (2019). Понятие искусственного интеллекта и юридическая ответственность за его работу. Право. *Журнал Высшей школы экономики*, *2*, 79–102. EDN: https://elibrary.ru/gqatho. DOI: https://doi.org/10.17-323/2072-8166.2019.2.79.102

Пибаев, И. А., Симонова, С. В. (2020). Алгоритмы в механизме реализации конституционных прав и свобод: вызовы цифровой эпохи. *Сравнительное конституционное обозрение*, *6*, 31–50. EDN: https://elibrary.ru/zmmnic. DOI: https://doi.org/10.21128/1812-7126-2020-6-31-50

Харитонова, Ю. С., Савина, В. С., Паньини, Ф. (2022). Гражданско-правовая ответственность при разработке и применении систем искусственного интеллекта и робототехники: основные подходы. *Вестник Пермского университета. Юридические науки*, *58*, 683–708. EDN: https://elibrary.ru/ppvmzr. DOI: https://doi.org/10.17072/1995-4190-2021-58-683-708

Чаннов, С. Е. (2022). Робот (система искусственного интеллекта) как субъект (квазисубъект) права. *Актуальные проблемы российского права*, *12*, 94–109. EDN: https://elibrary.ru/memsif. DOI: https://doi.org/10.17803/1994-1471.2022.145.12.094-109

Шахназаров, Б. А. (2022). Правовое регулирование отношений с использованием искусственного интеллекта. *Актуальные проблемы российского права*, *9*, 63–72. EDN: https://elibrary.ru/yownjo. DOI: https://doi.org/10.17803/1994-1471.2022.142.9.063-072

Bertolini, A. (2013). Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules. *Law, Innovation and Technology*, *5*. https://doi.org/10.5235/17579961.5.2.214

Bokovnya, A. Y. et al. (2020). Legal Approaches to Artificial Intelligence Concept and Essence Definition. *Revista San Gregorio*, *41*, 115–121. https://doi.org/10.36097/rsan.v1i41.1489

Duffy, S. H., & Hopkins, J. P. (2013). Sit, Stay, Drive: The Future of Autonomous Car Liability. *SMU Science & Technology Law Review*, *16*.

Leenes, R., & Lucivero, F. (2014). Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design. *Law, Innovation and Technology*, *6*(2), 194–222.

Llorca, D. F. (2023). Liability Regimes in the Age of AI: a Use-Case Driven Analysis of the Burden of Proof. *Journal of Artificial Intelligence Research*, *76*, 613–644. https://doi.org/10.48550/arXiv.2211.01817

Philipp, H. (2023). The European AI liability directives — Critique of a half-hearted approach and lessons for the future. *Computer Law & Security Review*, *51*, 1–42. https://doi.org/10.1016/j.clsr.2023.105871

Tianran, L. (2024). Research on Legal Responsibility Attribution for Autonomous Systems: An AI Governance Perspective. *Science of Law Journal*, *3*(7), 166–174. https://doi.org/10.23977/law.2024.030722

Сведения об авторе

Казанцев Дмитрий Александрович – кандидат юридических наук, член Совета по развитию закупок, Торгово-промышленная палата Российской Федерации
Адрес: 109012, Россия, г. Москва, ул. Ильинка, 6/1с1
E-mail: info@dkazantsev.ru
ORCID ID: https://orcid.org/0000-0003-2182-5776
РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=1149755

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.19.31 / Право на информацию
Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 22 мая 2025 г.
Дата одобрения после рецензирования – 4 июня 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.

Research article
UDC 34:004:340.1.721:004.8
EDN: https://elibrary.ru/bvlgsu
DOI: https://doi.org/10.21202/jdtl.2025.19

Techno-Ethical and Legal Aspects of Cyberbullying: Structural Analysis of Power Imbalances in the Digital Sphere

Fotios Spyropoulos

Philips University, Nicosia, Cyprus
Spyropoulos Law Firm, Athens, Greece

Keywords

algorithmic transparency,
artificial intelligence,
cyberbullying,
digital platforms,
digital security,
digital technologies,
ethics,
international cooperation,
law,
technological literacy

Abstract

Objective: to conceptualize cyberbullying from the viewpoint of law and technoethics; to analyze the power imbalance in the digital environment as a fundamental factor of causing harm online.

Methods: the work uses a conceptual and analytical methodology based on an interdisciplinary analysis of the theoretical provisions of law, technoethics, philosophy of technology, and social psychology. The methodological tools are complemented by constructing unique conceptual models through analyzing the structural factors of the digital space, developing causal relationships and creating a taxonomy of cyberbullying forms. Special attention is paid to the comparative analysis of regulatory approaches of different jurisdictions and the identification of gaps in existing legal norms.

Results: the research established that cyberbullying is a complex multilevel phenomenon that occurs at the intersection of the architectural features of digital platforms, the asymmetry of technological competencies between participants in interactions, and the systemic fragmentation of legislative regulation. It identified the critical gaps in key international legal instruments, manifested in the lack of unified definitions of cyberbullying, insufficiently elaborated mechanisms for cross-border cooperation, and irrelevant addressing of the digital environment specifics. The author analyzed the fundamental ethical issues related to automated content moderation based on machine learning algorithms, the distribution of responsibility between platforms, government regulators and individual users, and the contradictions between ensuring security and maintaining user autonomy. Four main types of power imbalances were identified: technological, informational, social, and institutional; each of them requires specific strategies to overcome.

Scientific novelty: for the first time, the article proposed a comprehensive approach to analyzing cyberbullying as a structurally determined abuse of digital power through the prism of technoethics. The developed conceptual models provide new tools for understanding the distributed nature of responsibility in the digital ecosystem and forming ethically sound prevention strategies. The author introduced a concept of information misuse as a central mechanism of systematic abuse of power in the digital environment.

Practical significance: the research is aimed at legal scholars, public officials, and digital platform developers. It offers practical solutions in the fields such as ethical audit of algorithms, creation of hybrid moderation systems involving artificial intelligence and humans, formation of international task forces, and development of human rights-based principles of digital literacy. The author's proposals may help to create a safer, more accountable and inclusive digital environment for all participants.

For citation

Spyropoulos, F. (2025). Techno-Ethical and Legal Aspects of Cyberbullying: Structural Analysis of Power Imbalances in the Digital Sphere. *Journal of Digital Technologies and Law, 3*(3), 472–496. https://doi.org/10.21202/jdtl.2025.19

Contents

Introduction

Cyberbullying, although increasingly prevalent, lacks a universally accepted definition both in Europe and internationally (Smith et al., 2013). According to UNICEF, it is

defined as "bullying with the use of digital technologies"[1]. Similarly, the European Commission[2] describes it as "repeated verbal or psychological harassment carried out by an individual or group, leveraging digital platforms to disseminate harmful content, such as abusive messages or embarrassing photos, with the aim of distressing or humiliating victims".

The United Nations recognizes cyberbullying as a form of online violence characterized by an imbalance of power, anonymity, and a broad audience. Unlike traditional bullying, a single harmful act online can constitute cyberbullying due to the permanent and far-reaching nature of digital content[3]. This evolving definition reflects the unique risks posed by digital platforms, including constant accessibility and the replication of harmful material, which exacerbate the victim's vulnerability (Langos, 2012; Menesini et al., 2012; Slonje & Smith, 2008).

Scholarly definitions often build upon Olweus's (1993) traditional bullying framework, emphasizing the use of digital tools, intent to harm, and repeated actions (Englander et al., 2017; Mouzaki, 2010; Juvonen & Gross, 2008). Central to cyberbullying is the aggressor's exploitation of technological advantages to target victims who may lack the means to defend themselves, as well as the anonymity and publicity provided by digital platforms (Kowalski, 2018; Smith et al., 2008; Nocentini et al., 2010; Hinduja & Patchin, 2006). Li's (2007) metaphor, "New bottle but old wine," aptly captures the way cyberbullying mirrors traditional bullying while incorporating the distinct features of digital technology.

Artificial intelligence (AI) introduces additional complexities to this landscape, serving both as a tool for addressing cyberbullying and a potential challenge. AI-powered systems are increasingly used to detect harmful content, moderate interactions, and prevent the spread of abusive material. However, these systems often face limitations, such as difficulties in understanding context, cultural nuances, or distinguishing harmful intent from satire or criticism. Additionally, aggressors have begun exploiting AI tools, such as deepfakes or automated bots, to amplify harm, manipulate content, or target victims on a larger scale. These developments underscore the need for robust, transparent, and ethical AI systems to counteract cyberbullying effectively (Hasan et al., 2023; Raj et al., 2022).

The psychological and social consequences of cyberbullying, particularly among children and adolescents, are profound, ranging from mental health issues to damaged

1 UNICEF, n.d. Cyberbullying: What is it and how to stop it. https://clck.ru/3NQaBe

2 European Commission. (2009). Safer Internet Programme: Protecting children online. https://clck.ru/3NQaRi

3 United Nations. (2016). Ending the Torment: Tackling Bullying from Schoolyard to Cyberspace. https://clck.ru/3NQaWz; United Nations. (2016). Convention on the Rights of the Child: General Comment No. 20 (2016) on the Implementation of the Rights of the Child during Adolescence (CRC/C/GC/20). https://clck.ru/3NQaZ3

relationships (Campbell & Bauman, 2018). Additionally, specific harmful behaviors, such as the unauthorized dissemination of explicit images ("sexting"), further highlight the dangers posed by cyberbullying (Katerelos et al., 2011; Chakraborty et al., 2021).

Despite these insights, the absence of a standardized definition continues to hinder global efforts to combat cyberbullying effectively. Addressing this gap requires comprehensive approaches, including educational campaigns, digital literacy programs, stricter regulations, and international collaboration. Recognizing the unique dynamics of cyberbullying, including the evolving role of AI, is essential for developing interventions that create safer and more equitable digital spaces.

1. The Misuse of Information as a Systematic Abuse of Power in Cyberbullying

The misuse of information within digital environments has become a defining feature of cyberbullying, representing a systematic abuse of power. In these contexts, aggressors exploit technological tools to manipulate, control, and harm others, capitalizing on the unique affordances of the internet. Unlike traditional forms of bullying, the digital sphere enables perpetrators to transcend physical boundaries, leveraging the scalability of online platforms, anonymity, and the permanence of digital content to amplify their actions (Courakis, 2005; Lazos, 2001; Furnell, 2006).

Central to this phenomenon is the Enhanced Cyberbullying Power Imbalance Model (Figure 1), which provides a framework for understanding the dynamics of power in digital bullying. The model highlights key factors that facilitate harm, including the aggressor's ability to manipulate information, exploit anonymity, and reach wide audiences. These elements not only empower the aggressor but also exacerbate the vulnerability of victims by creating a sustained and pervasive impact.

A key addition to the model is the concept of information misuse, which represents acts such as unauthorized access, manipulation, or dissemination of private content. Cyberbullies frequently weaponize information to undermine their victims' psychological well-being and social standing. Examples include the sharing of sensitive photos, creation of fake profiles, or spreading defamatory material. These actions exemplify how the digital environment reshapes traditional power dynamics, allowing aggressors to assert dominance while evading accountability (Spyropoulos, 2011; Katerelos et al., 2011).

The systematic abuse of information is further reinforced by disparities in technological knowledge and familiarity with technology. Aggressors often possess advanced skills that allow them to exploit digital tools with greater precision, while victims, particularly those with limited digital literacy, are left unable to respond effectively. This knowledge gap deepens the power imbalance, making victims feel isolated and disempowered (Vandebosch & Van Cleemput, 2008; Ybarra & Mitchell, 2004a, 2004b).

These dynamics are supported by Gibson's (2014) Theory of Affordances, which explains how digital tools shape user behavior. In cyberbullying, technological affordances such as anonymity and the scalability of harm allow aggressors to act with impunity, amplifying the psychological and social damage inflicted on victims (Topcu-Uzer & Tanrıkulu, 2018). For instance, the widespread availability of platforms like social media enables perpetrators to reach larger audiences while shielding themselves from detection.

At a macro level, this power imbalance is linked to broader structural factors. Socio-economic disparities often determine access to technological expertise, reinforcing systemic inequalities. Those in privileged positions are more likely to acquire advanced knowledge and resources, enabling them to manipulate information as a tool of control and dominance. This dynamic is mirrored in larger phenomena, such as political cyberbullying, cyberterrorism, and information warfare, where control over technology and data is central to power struggles (Millard, 2009; Zannis, 2005; Bosworth et al., 1999).

The "Enhanced Cyberbullying Power Imbalance Model" underscores the interplay between these individual and systemic factors. It illustrates how aggressors leverage technological tools and knowledge gaps to consolidate their dominance, making interventions particularly challenging. The model calls for targeted strategies that address these imbalances at both individual and structural levels (Figure 1).

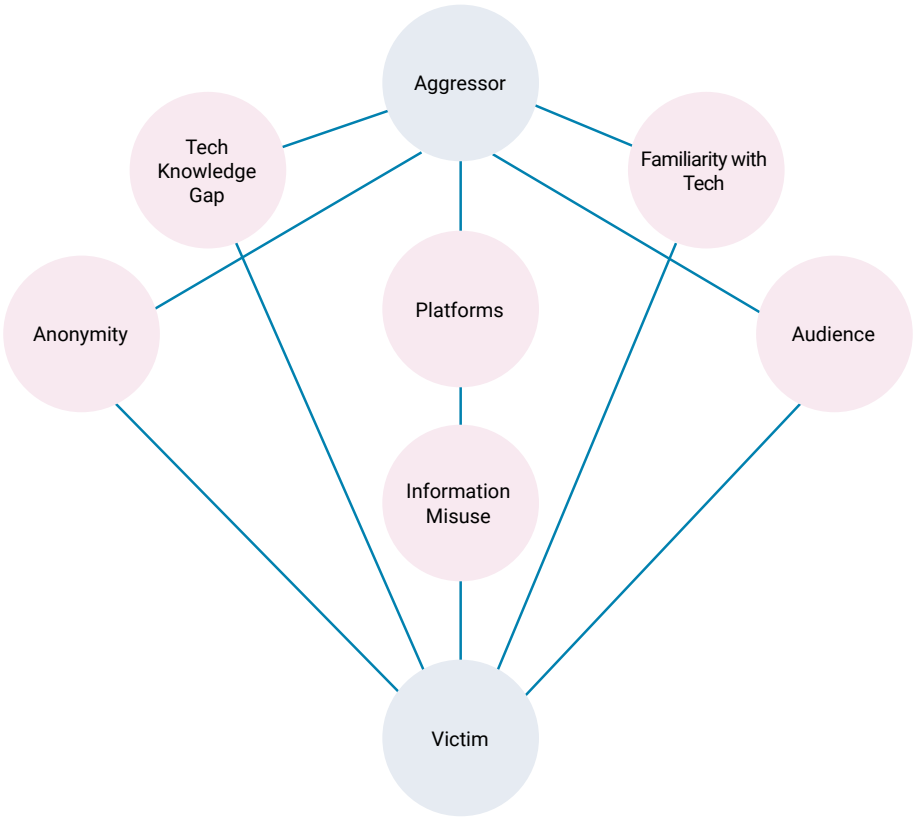

Figure 1. Enhanced Cyberbullying Power Imbalance Model, incorporating the role of information misuse and technological literacy in cyberbullying dynamics

So, the systematic misuse of information is at the heart of the power imbalance inherent in cyberbullying. This phenomenon is driven not only by individual behavior but also by technological and systemic factors that reshape traditional notions of harm and control. Addressing this issue requires a multifaceted approach. Interventions must focus on bridging gaps in digital literacy, empowering victims with resilience, and fostering greater accountability among platforms. Moreover, promoting equitable access to technology and implementing robust ethical and legal safeguards are critical to rebalancing the dynamics of power in the digital sphere. By integrating these strategies, stakeholders can work toward a safer, more equitable, and inclusive digital environment.

2. Cyberbullying and Technoethics: A Technoethical Flow Analysis

Cyberbullying constitutes not only a sociotechnical phenomenon but also a fundamental ethical challenge in the digital age. As Luppicini (2018) observes, technoethics offers a conceptual lens through which we can examine the intersection of technology and human values, shedding light on the misuse of digital platforms for aggressive or harmful behaviors. Cyberbullying – defined as the intentional and repeated harm inflicted through digital devices – amplifies these concerns, as aggressors exploit affordances such as anonymity, virality, and permanence to target victims (Langos, 2012; Menesini et al., 2013).

The technoethical implications are manifold. Firstly, cyberbullying undermines the principle of digital dignity, defined as the right of individuals to exist in online spaces free from humiliation and harm (Verbeek, 2011). Platforms often fail to intervene effectively, despite possessing technological capabilities to moderate or flag harmful content[4]. This reveals a gap between technological potential and ethical implementation. As Moor (2005) argues, emerging technologies require evolving ethical frameworks – ones that anticipate misuse and promote societal well-being.

Secondly, cyberbullying highlights power imbalances embedded in digital infrastructures. The capacity to harm is not evenly distributed; aggressors often possess greater technological fluency, while victims may lack digital literacy or access to effective reporting mechanisms (Spyropoulos, 2011; Katerelos et al., 2011). These asymmetries, visualized in the Technoethical Responsibilities Flowchart (Figure 2), offer a structured mapping of responsibility across multiple actors.

[4] Hinduja, S., & Patchin, J. W. (2014). Cyberbullying: Identification, Prevention and Response. Cyberbullying Research Center. https://clck.ru/3NQcNU

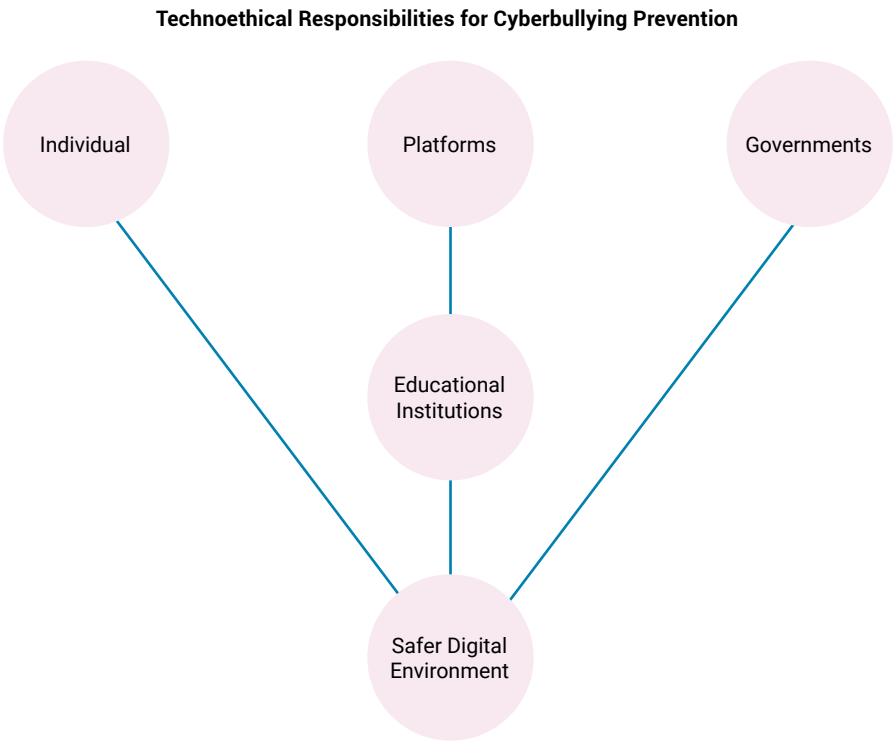

Figure 2. Technoethical Responsibilities Flowchart, highlighting the collaborative roles of individuals, platforms, governments, and educational institutions in addressing cyberbullying

This conceptual tool identifies four primary layers of responsibility:

1. Design-Level Responsibility: At the foundational level lie system designers and developers. Their choices in platform architecture, moderation features, and affordances shape user interactions. If anonymity is permitted without safeguards, or if virality is incentivized without accountability, then cyberbullying becomes more likely (Capurro, 2009; Tavani, 2011).

2. Operational Responsibility: Platform operators and content moderators are ethically obligated to monitor, detect, and remove harmful content, while preserving freedom of expression. Failure to act promptly or transparently exacerbates victimization (Zuboff, 2019).

3. User-Level Responsibility: Ethical conduct is not the sole burden of institutions. Users must exercise empathy, restraint, and digital citizenship. Educational programs targeting youth can instill these technoethical values (Chen, 2017; Ortega-Ruiz et al., 2012).

4. Regulatory-Level Responsibility: While not extensively addressed here, frameworks such as the GDPR enforce transparency, accountability, and user rights, embodying technoethical norms in legal terms[5]. These frameworks are examined in greater detail in subsequent sections.

[5] European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88. https://clck.ru/3NQbz7

The flowchart illustrates that technoethical responsibility in cyberbullying contexts is not linear but distributed and interdependent. No single stakeholder can resolve the issue in isolation. The strength of technoethical analysis lies in emphasizing this interconnectedness.

Moreover, automation and AI-based detection tools introduce new layers of complexity. While such systems may identify patterns of abuse, they can also reproduce bias or fail to grasp contextual nuance (Ioannou et al., 2018). Therefore, transparency and human oversight remain indispensable (Tavani, 2011).

In sum, cyberbullying – when viewed through a technoethical framework – reveals the necessity for shared responsibility, ethical system design, proactive intervention, and the cultivation of respect for digital dignity across all levels of the sociotechnical ecosystem.

3. The Nexus Between Technoethics, Cyberbullying, and Its Prevention

Technoethics provides a critical framework for examining the ethical dimensions of cyberbullying, offering a robust foundation for the development of policies and strategies to mitigate its impact. By focusing on the shared responsibilities of individual users, technological platforms, and regulatory bodies, technoethics advocates for a safer, more equitable digital environment. The prevention and management of cyberbullying require a comprehensive approach that goes beyond the application of technological solutions or isolated legislative measures. Importantly, most of these prevention frameworks are deeply rooted in psychological science, emphasizing emotional resilience, empathy-building, and behavioral awareness. Thus, effective intervention must integrate ethical principles, educational initiatives, and innovative psychosocial strategies to ensure that technological progress aligns with human rights and societal well-being (Hinduja & Patchin, 2009).

The General Data Protection Regulation (GDPR) constitutes one of the most significant legal instruments for safeguarding personal data within the European Union[6], particularly in contexts of online harm. By enshrining principles such as data minimization and the right to erasure, it empowers victims of cyberbullying to reclaim control over their digital presence and seek redress against the misuse of personal information. More than a regulatory mechanism, the GDPR embodies core technoethical values – transparency, accountability, and autonomy – transforming digital rights into enforceable ethical protections. In doing so, it contributes to a more respectful and human-centered digital environment[7].

---

6 Ibid.

7 European Parliament and Council of the European Union. (2022). Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148. Official Journal of the European Union, L333, 80–137. https://clck.ru/3QGqvZ

Equally critical is the Artificial Intelligence Act (AI Act), introduced by the European Union as the first comprehensive global framework to regulate AI technologies (Regulation (EU) 2024/1689). The AI Act employs a risk-based approach to classify AI systems, ranging from minimal to high risk. Tools used for content moderation and the detection of harmful online behaviors, such as cyberbullying, are typically categorized as high-risk systems, necessitating strict compliance with transparency, fairness, and oversight requirements. These obligations ensure that AI systems employed on social media platforms or in educational settings are accurate, impartial, and supervised by human operators. Furthermore, the AI Act prohibits manipulative technologies that could influence user behavior in harmful ways, reflecting its commitment to safeguarding user rights while promoting ethical innovation.

At the international level, the United Nations Convention against Cybercrime, adopted in 2024, represents a significant milestone in addressing ICT-facilitated crimes. While the convention provides a comprehensive framework for combating cybercrime, including illegal system access, data breaches, and online fraud, it notably excludes explicit provisions for cyberbullying. This omission highlights the ongoing challenges in achieving a unified global approach to addressing this widespread issue. Despite this gap, several provisions within the convention have indirect relevance to cyberbullying. For instance, Article 14 addresses online child sexual abuse, and Article 16 targets the unauthorized dissemination of intimate images, behaviors often associated with cyberbullying. Furthermore, Article 18 establishes the liability of platforms that facilitate harmful activities, while Article 34 includes measures for victim assistance and protection[8].

The absence of explicit references to cyberbullying within the United Nations Convention against Cybercrime[9]. Underscores the need for further amendments or supplementary protocols. A harmonized international framework addressing cyberbullying is essential to bridge this gap, especially given the transnational nature of many incidents. Victims of cyberbullying often face significant barriers to justice when perpetrators operate across jurisdictions. Establishing a universal definition of cyberbullying would provide a foundation for coordinated efforts, enabling clearer pathways for legal action and victim support. In addition, aligning the convention with technoethical principles could enhance its applicability, ensuring that regulatory measures reflect the ethical imperatives of fairness, accountability, and user protection.

---

8 United Nations General Assembly. (2024). United Nations Convention against Cybercrime: Strengthening international cooperation for combating crimes committed via ICT systems and evidence sharing. https://clck.ru/3NQc2k

9 Ibid.

Despite these limitations, the convention serves as an important step toward global collaboration on cyber-related crimes. Its emphasis on cross-border cooperation and its focus on shared responsibilities among member states create a framework that could be adapted to address the unique challenges of cyberbullying. As digital platforms continue to evolve, the inclusion of cyberbullying-specific provisions in future revisions of the convention would strengthen its relevance and effectiveness.

By integrating technoethical principles into regulatory innovations such as the GDPR, AI Act, and United Nations Convention against Cybercrime (United Nations General Assembly, 2024), stakeholders can create a digital ecosystem that prioritizes safety, accountability, and inclusivity. Collaboration between governments, platforms, and civil society organizations is essential for developing cohesive strategies that align with ethical imperatives. For instance, partnerships with researchers and policymakers could facilitate the design of advanced tools for harm detection and mitigation. Furthermore, promoting best practices, such as transparency reports and algorithm audits, can ensure that platforms operate in ways that are both ethical and effective.

4. Practical and Interdisciplinary Recommendations

4.1. Technological Tools

The role of technology in preventing and managing cyberbullying has become increasingly significant, with modern platforms adopting innovative solutions to address harmful online behaviors. Major platforms, employ advanced artificial intelligence (AI) tools to detect and remove harmful content. According to Meier et al. (2016), AI systems are particularly effective in real-time monitoring, identifying suspicious behavior, and mitigating the spread of offensive posts. For instance, automated reporting systems can significantly reduce victims' exposure to abusive comments by flagging and removing harmful content promptly.

However, while AI offers powerful tools for moderating harmful content, it is not without limitations. As Zuckerberg[10] noted, AI systems struggle to understand complex contexts and cultural nuances, making them less effective in situations requiring subtle judgment. A hybrid approach that combines AI technology with human oversight is essential to ensure accuracy and fairness in content moderation. This perspective aligns with Zuckerberg's[11] emphasis on collective user involvement through tools like Community Notes, which

[10] Zuckerberg*, M. (2024). It's time to get back to our roots around free expression. Facebook Watch. https://clck.ru/3NQcAk (* A co-founder of Meta , a company banned in the Russian Federation and included in the list of extremist organizations.)

[11] A co-founder of Meta, a company banned in the Russian Federation and included in the list of extremist organizations.

empower users to contribute context and collaboratively address misleading information or harmful content.

From a technoethical standpoint, the deployment of surveillance and monitoring technologies in combating cyberbullying must balance safety with privacy. Floridi (2014) argues that transparency, respect for individual rights, and robust data protection are vital to avoiding oppressive practices. The ethical use of technology requires systems that prioritize human dignity while delivering effective solutions to reduce online harm.

In addition to advancing technological tools, the need for multi-stakeholder collaboration is paramount. Governments, civil society, and technology platforms must co-develop content moderation protocols that balance effectiveness with fairness. These collaborations should focus on creating transparent systems that prioritize harm prevention while safeguarding user rights. By pooling expertise and resources, stakeholders can develop adaptable and culturally sensitive solutions to address the complexities of cyberbullying across diverse digital environments.

Partnerships with educational institutions further enhance the efficacy of prevention strategies. Schools and universities can integrate digital ethics and resilience training into their curricula, equipping students with the skills needed to navigate digital spaces safely and responsibly. These programs should emphasize critical thinking, empathy, and digital literacy, fostering a culture of respect and accountability among younger generations. Collaborative initiatives between educators, policymakers, and platforms can create comprehensive educational frameworks that address the root causes of cyberbullying while promoting ethical digital behavior.

By combining technological innovation with ethical principles and interdisciplinary collaboration, stakeholders can address the complexities of cyberbullying effectively. This approach not only mitigates harm but also upholds the broader principles of fairness, accountability, and respect for human dignity, ensuring that technological progress aligns with societal values.

4.2. Digital Citizenship and Promoting Responsible Behavior

Technoethics highlights the vital importance of fostering digital citizenship, advocating for the cultivation of responsible and respectful behavior in online interactions. Education in digital citizenship encompasses essential principles such as demonstrating respect for others, upholding a sense of accountability in virtual environments, and refraining from harmful behaviors, including cyberbullying. The promotion of these values is integral to the prevention of online misconduct, as it strengthens social

responsibility and fosters a culture of respect within the broader digital and educational environment[12].

4.3. Protection of Victims

The effective protection of victims of cyberbullying requires the establishment of structured and victim-centered reporting frameworks. Such frameworks should enable secure and confidential reporting of incidents, ensuring victims feel safe, supported, and empowered throughout the process. Transparency and trust are critical in managing complaints effectively, fostering an environment where individuals are confident in seeking help. From a technoethical perspective, technological tools must prioritize victim welfare, integrating features that enhance safety and provide accessible avenues for reporting and support.

Confidential and secure reporting channels are essential for achieving these objectives. Educational institutions and communication platforms can play a pivotal role by implementing systems that allow students and users to report bullying or harassment without fear of retaliation or exposure. These channels should be designed to offer anonymity and user protection while ensuring swift and effective resolution. For instance, Ioannou et al. (2018) highlight the importance of integrating such tools into school networks and digital platforms, reinforcing the ethics of care in both educational and online environments.

To complement these frameworks, governments should consider funding dedicated cyberbullying hotlines. These hotlines would provide victims with immediate access to psychological support and guidance, connecting them with trained professionals who can offer advice and assistance tailored to their needs. Such initiatives can bridge critical gaps in victim support, particularly in cases where individuals may lack access to other resources. By addressing the psychological and emotional dimensions of cyberbullying, these hotlines contribute to the holistic protection of victims and the promotion of mental well-being.

In addition to reactive measures, a proactive approach to victim protection can be achieved through the development of a Digital Resilience Index. This tool would assess the capacity of vulnerable groups, such as children and adolescents, to navigate cyber risks safely and effectively. The index could evaluate factors such as digital literacy, emotional resilience, and access to support systems, providing valuable insights into the specific needs of at-risk populations. By identifying areas for improvement, the Digital Resilience

[12] Bynum, T. W. (2008). Computer and Information Ethics. In The Stanford Encyclopedia of Philosophy. https://clck.ru/3NQcEy

Index can guide targeted interventions, including education, awareness campaigns, and policy adjustments.

Such victim-centered strategies not only mitigate the immediate impacts of cyberbullying but also foster a culture of empathy and support within digital and educational spaces. Tavani (2011) emphasizes that prioritizing victim welfare and cultivating a sense of safety are integral to building resilience and empowering individuals to reclaim control over their digital experiences. Collaborative efforts among governments, platforms, and civil society organizations are crucial in ensuring these frameworks are both effective and widely accessible.

By integrating secure reporting channels, government-funded support services, and tools like the Digital Resilience Index, stakeholders can create a comprehensive system for victim protection. These measures, grounded in technoethical principles, address the multifaceted challenges of cyberbullying while promoting a more inclusive and supportive digital environment.

4.4. Ethical Dilemmas

The use of technology to prevent cyberbullying, while yielding significant benefits, raises complex ethical and legal dilemmas that demand careful examination. One key issue is the protection of privacy, particularly in the context of data surveillance practices employed by social media platforms. These practices aim to identify suspicious behavior and prevent cyberbullying but carry the inherent risk of abuse of power. The collection and analysis of large datasets, often conducted without explicit user consent, can lead to potential violations of privacy rights and personal autonomy, as protected under frameworks like the General Data Protection Regulation (GDPR). This raises critical questions about the extent to which privacy can be compromised to ensure online safety.

Another pressing concern involves the deployment of content moderation algorithms. While these tools are designed to detect and remove harmful content, they often lack the nuance to distinguish between hate speech and lawful expressions of sarcasm or criticism. Such limitations can result in unintended censorship, restricting freedom of expression, a right enshrined in international agreements such as the European Convention on Human Rights (Article 10 ECHR)[13]. This scenario undermines the democratic exchange of ideas and highlights the need for safeguards to prevent overreach.

These dilemmas underscore the necessity of balancing the rights to privacy and freedom of expression with the imperative to protect user safety and dignity. Technoethics

[13] Council of Europe. (1950). European Convention on Human Rights, Article 10: Freedom of Expression. https://clck.ru/3NQngR

offers a framework for addressing these challenges by promoting transparency in surveillance practices and ensuring accountability in the design and implementation of content moderation algorithms. Establishing such frameworks can prevent abuses, foster trust in digital platforms, and ensure that technological interventions respect fundamental rights while enhancing online safety.

4.5. Strengthening Collaboration and Data-Driven Insights

The transnational nature of cyberbullying necessitates a unified and collaborative global response. Due to its ability to transcend national borders, cyberbullying challenges traditional jurisdictional boundaries and requires harmonized legal frameworks to ensure consistent protections for victims worldwide. The decentralized structure of the internet often allows perpetrators to exploit disparities in national laws, making international cooperation imperative. A global approach grounded in shared principles of justice, accountability, and human rights is necessary to address these challenges effectively.

A key recommendation is the establishment of an International Cyberbullying Prevention Taskforce, which would facilitate cross-border investigations, enable information sharing, and support coordinated law enforcement efforts. This body could bridge jurisdictional gaps by developing internationally recognized protocols for handling cyberbullying cases and ensuring that perpetrators face accountability regardless of geographic location. Additionally, it could work toward the harmonization of national legal frameworks with international standards, reducing policy fragmentation and ensuring victims receive equitable legal protections.

At the policy level, several national initiatives demonstrate promising approaches to addressing cyberbullying. Countries have implemented diverse strategies, ranging from digital literacy campaigns and technological tools to specialized legal frameworks that recognize the unique harm inflicted by online aggression. For example, Greece has introduced the «Safe Youth» digital application[14], an innovative tool designed to support minors aged 12 and older in addressing online threats. This application provides direct communication with emergency services, discreet emergency notifications for real-time threats, and a secure system for submitting abuse reports through the Unified Digital Portal of Public Administration. By prioritizing accessibility, confidentiality, and

[14] Ministry of Citizen Protection, Hellenic Police and Vodafone Foundation. (2024). SAFE.YOUth: Digital Application for the Protection of Minors. https://clck.ru/3NQcsT

immediacy, the Safe Youth initiative represents a technoethical approach to digital safety, empowering minors to navigate crises safely. However, ensuring equitable access to this tool – particularly for marginalized populations or those with limited technological literacy – remains a critical challenge.

Despite the progress of national initiatives like Safe Youth, a significant challenge in existing legal frameworks is their adaptability to emerging technological threats. Many laws focus on traditional online harassment mechanisms but fail to account for algorithm-driven amplification of harmful content, misuse of artificial intelligence (AI), and the role of digital anonymity in perpetuating abuse. Future regulatory frameworks must address these evolving dynamics, particularly regarding social media algorithms, gaming platforms, and AI-generated content (e.g., deepfakes). Expanding legal protections to cover these areas is crucial to maintaining the relevance and effectiveness of legislative measures.

From a criminological perspective, cyberbullying reflects an imbalance of power, where perpetrators leverage anonymity and the expansive reach of digital platforms to act with impunity. While legal accountability mechanisms provide recourse for victims, long-term cultural and systemic shifts are needed to address the broader social structures that enable digital aggression. Integrating restorative justice principles (Guardabassi & Nicolini, 2024; Duncan, 2016), rehabilitation programs for offenders (Othman et al., 2024), and sustained victim support systems can contribute to a more holistic approach to prevention and intervention (Vandebosch, 2019).

In parallel with legal and institutional responses, the integration of data-driven insights is critical for understanding cyberbullying trends and designing effective interventions. Conducting meta-analyses on the prevalence of cyberbullying and examining platform-specific risk factors can help policymakers and researchers develop targeted prevention strategies (Sathya & Fernandez, 2024; Kim et al., 2021). For instance, studying how algorithms amplify harmful content or how anonymous features facilitate harassment can provide valuable insights into mitigating the negative impact of these technologies (Johora et al., 2024; Meier et al., 2016).

Furthermore, anonymized data-sharing agreements between digital platforms and research institutions represent a promising avenue for combating cyberbullying. Social media platforms collect vast amounts of behavioral data that, when responsibly anonymized, could be leveraged to identify patterns of abuse and develop proactive solutions. Transparent agreements that balance privacy rights with research accessibility can foster innovation while protecting individual users (Floridi, 2014).

By strengthening international collaboration, refining national legal frameworks, and leveraging data-driven insights, stakeholders can build a cohesive, global strategy to combat cyberbullying. Initiatives like Greece's Safe Youth demonstrate the potential of digital tools in mitigating online threats, but they must be complemented by comprehensive international efforts. Integrating technoethical principles – which prioritize accountability, digital rights, and equitable access to protection mechanisms – will be key in fostering a safer, more inclusive, and resilient digital environment for future generations.

Conclusions

Cyberbullying, as a pervasive form of digital aggression, exemplifies the misuse of technology to harm, intimidate, or humiliate individuals. Addressing this multifaceted issue demands a holistic approach that integrates prevention, legal mechanisms, technological innovation, and ethical considerations. Central to these efforts is the recognition of structural inequities, such as disparities in technological access and literacy, which amplify vulnerabilities and perpetuate power imbalances in digital interactions (Lazos, 2001). Empowering users with knowledge and fostering an ethical culture of information usage remain crucial steps toward promoting responsible engagement in digital environments (Tsouramanis, 2005).

Policy recommendations must emphasize adaptive and forward-looking legal frameworks at both national and international levels. Governments should collaborate to harmonize laws addressing cyberbullying, ensuring consistency and accountability across jurisdictions. The establishment of an International Cyberbullying Prevention Taskforce, as well as a Global Cyberbullying Prevention Treaty, would strengthen cross-border investigations and provide unified standards for combating this issue. These frameworks must be dynamic, incorporating emerging technologies such as artificial intelligence, blockchain, and algorithmic transparency, which play a dual role in enabling and preventing harm.

Collaboration among stakeholders is essential for creating a comprehensive response. Technology platforms bear a critical responsibility to design and implement robust content moderation systems, transparency mechanisms, and user protection measures. Governments must legislate adaptive laws that balance individual freedoms with protections against online harm. Civil society organizations and educational institutions also play pivotal roles in shaping ethical norms, promoting awareness, and providing support for victims.

Prevention remains the cornerstone of addressing cyberbullying. Educational institutions should integrate digital literacy, resilience-building, and ethical technology use into curricula, empowering students to navigate online spaces responsibly.

Government-funded awareness campaigns and parental engagement initiatives should complement these efforts, fostering a culture of empathy and accountability. Moreover, bridging the digital divide and promoting equitable access to technology are critical for addressing systemic inequities that exacerbate cyberbullying. Ensuring all users – particularly vulnerable groups – have the tools and knowledge to engage safely in digital spaces aligns with the ethical imperatives of inclusivity and social justice.

Ethical dilemmas persist in the implementation of technological tools, such as AI-driven content moderation and surveillance technologies, which must balance effectiveness with respect for privacy and human dignity. The principles of technoethics – responsibility, accountability, and inclusivity – must guide these innovations to ensure they align with societal well-being rather than perpetuate harm.

In conclusion, combating cyberbullying requires a multidisciplinary, collaborative approach that integrates policy innovation, stakeholder engagement, and ethical foresight. By addressing root causes, promoting equitable access, and fostering a culture of responsibility, society can mitigate the harms of cyberbullying. Such efforts ensure that technological advancements contribute to the creation of a safer, more inclusive digital ecosystem, where the rights, dignity, and well-being of all users are prioritized.

References

Bosworth, K., Espelage, D. L., & Simon, T. R. (1999). Factors associated with bullying behavior in middle school students. *The Journal of Early Adolescence*, *19*(3), 341–362. https://doi.org/10.1177/0272431699019003003

Chakraborty, S., Bhattacherjee, A., & Onuchowska, A. (2021). Cyberbullying: A review of the literature. SSRN. http://dx.doi.org/10.2139/ssrn.3799920

Campbell, M., & Bauman, S. (2018). Cyberbullying: Definition, consequences, prevalence. In M. Campbell & S. Bauman (Eds.), *Reducing Cyberbullying in Schools* (pp. 3–16). Academic Press. https://doi.org/10.1016/B978-0-12-811423-0.00001-8

Capurro, R. (2009). Digital ethics. *The Information Society*, *25*(3), 183–186. https://doi.org/10.1080/01972240902848902

Chen, C. W. Y. (2017). "Think before you type": The effectiveness of implementing an anti-cyberbullying project in an EFL classroom. *Pedagogies*, *13*(1), 1–18. https://doi.org/10.1080/1554480X.2017.1363046

Courakis, N. (2005). *Criminological horizons. Vol. II: Pragmatic approach and individual issues*. 2nd ed. Athens – Komotini: Ant. N. Sakkoula (In Greek).

Duncan, S. H. (2016). Cyberbullying and restorative justice. In R. Navarro, S. Yubero & E. Larrañaga (Eds.). *Cyberbullying Across the Globe* (pp. 239–257). Cham: Springer International Publishing.

Englander, E., Donnerstein, E., Kowalski, R., Lin, C. A., & Parti, K. (2017). Defining cyberbullying. *Pediatrics*, *140*(Supplement_2), S148–S151. https://doi.org/10.1542/peds.2016-1758u

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Furnell, S. M. (2006). *Computer Insecurity: Risking the System*. Oxford: Chandos Publishing.

Gibson, J. J. (2014). The theory of affordances (originally published 1979). In G. Bridge & S. Watson (Eds.). *The People, Place, and Space Reader* (pp. 56–60). London: Routledge.

Guardabassi, V., & Nicolini, P. (2024). Adolescence and cyberbullying: a restorative approach. In *INTED2024 Proceedings* (pp. 4562–4570). IATED. https://doi.org/10.21125/inted.2024.1183

Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. (2023). A review on deep-learning-based cyberbullying detection. *Future Internet*, *15*(5), 179. https://doi.org/10.3390/fi15050179

Hinduja, S., & Patchin, J. W. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice*, *4*(2), 148–169. https://doi.org/10.1177/1541204006286288

Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, CA: Sage Publications.

Ioannou, A., Blackburn, J., Stringhini, G., De Cristofaro, E., Kourtellis, N., & Sirivianos, M. (2018). From risk factors to detection and intervention: a practical proposal for future work on cyberbullying. *Behaviour & Information Technology*, *37*, 258–266. https://doi.org/10.1080/0144929X.2018.1432688

Johora, F. T., Khan, M. S. I., Kanon, E., Rony, M. A. T., Zubair, M., & Sarker, I. H. (2024). A data-driven predictive analysis on cyber security threats with key risk factors. *arXiv preprint arXiv:2404.00068*.

Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? Bullying experiences in cyberspace. *Journal of School Health*, *78*(9), 496–505. https://doi.org/10.1111/j.1746-1561.2008.00335.x

Katerelos, I., Tsekeris, C., Lavdas, M., & Demetriou, K. (2011). A psychosociological approach to Internet use and mass online role-playing games. In K. Siomos & G. Floros (Eds.), *Research, prevention, management of risks in Internet use*. Larissa: Hellenic Society for the Study of Internet Addiction Disorder. (In Greek).

Kim, S., Razi, A., Stringhini, G., Wisniewski, P. J. and De Choudhury, M. (2021). A human-centered systematic literature review of cyberbullying detection algorithms. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2). https://doi.org/10.1145/3476066

Kowalski, R. (2018). Cyberbullying. In J. P. Forgas, R. F. Baumeister & T. F. Denson (Eds.), *The Routledge International Handbook of Human Aggression* (pp. 131–142). London: Routledge. https://doi.org/10.4324/9781315618777-11

Langos, C. (2012). Cyberbullying: The challenge to define. *Cyberpsychology, Behavior, and Social Networking*, *15*(6), 285–289. https://doi.org/10.1089/cyber.2011.0588

Lazos, G. (2001). *Information Technology and Crime*. Athens: Nomiki Bibliothiki. (In Greek).

Li, Q. (2007). New bottle but old wine: A research of cyberbullying in schools. *Computers in Human Behavior*, *23*(4), 1777–1791. https://doi.org/10.1016/j.chb.2005.10.005

Luppicini, R. (2018). *The Changing Scope of Technoethics in Contemporary Society*. Hershey, PA: Information Science Reference. https://doi.org/10.4018/978-1-5225-5094-5

Meier, A., Reinecke, L., & Meltzer, C. E. (2016). "Facebocrastination"? Predictors of using Facebook* for procrastination and its effects on students' well-being. *Computers in Human Behavior*, *64*, 65–76. https://doi.org/10.1016/j.chb.2016.06.011

Menesini, E., Nocentini, A., & Palladino, B. E. (2012). Empowering students against bullying and cyberbullying: Evaluation of an Italian peer-led model. *International Journal of Conflict and Violence*, *6*(2), 313–320. https://doi.org/10.4119/ijcv-2922

Menesini, E., Nocentini, A., & Camodeca, M. (2013). Morality, values, traditional bullying, and cyberbullying in adolescence. *British Journal of Developmental Psychology*, *31*(1), 1–14. https://doi.org/10.1111/j.2044-835X.2011.02066.x

Millard, G. (2009). Stephen Harper and the politics of the bully. *Dalhousie Review*, *89*(3), 329–336.

Moor, J. H. (2005). Why we need better ethics for emerging technologies. *Ethics and Information Technology*, *7*(3), 111–119. https://doi.org/10.1007/s10676-006-0008-0

Mouzaki, D. (2010). International scientific conference on: "Dealing with cyberbullying from a legal perspective". *The Art of Crime*, *15*. (In Greek).

Nocentini, A., Calmaestra, J., Schultze-Krumbholz, A., Scheithauer, H., Ortega, R., & Menesini, E. (2010). Cyberbullying: Labels, behaviours and definition in three European countries. *Australian Journal of Guidance & Counselling*, *20*(2), 129–142. https://doi.org/10.1375/ajgc.20.2.129

Olweus, D. (1993). *Bullying at School: What We Know and What We Can Do*. Oxford, UK: Blackwell.

Ortega-Ruiz, R., Del Rey, R., & Casas, J. A. (2012). Knowing, building and living together on internet and social networks: the ConRed cyberbullying prevention program. *International Journal of Conflict and Violence*, *6*(2), 303–313. https://doi.org/10.4119/ijcv-2921

Othman, S. N., Alziboon, M. F., Dawood, M., Sachet, S. J., & Moroz, I. (2024). New rehabilitation against electronic crimes by young people. *Encuentros: Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico*, *22*, 363–385. https://doi.org/10.5281/zenodo.13732745

Raj, M., Singh, S., Solanki, K., & Selvanambi, R. (2022). An application to detect cyberbullying using machine learning and deep learning techniques. *SN Computer Science*, *3*(5), 401. https://doi.org/10.1007/s42979-022-01266-w

* The social network belongs to Meta, which is recognized as an extremist organization, its functioning is prohibited in the territory of the Russian Federation.

Sathya, J., & Fernandez, F. M. H. (2024). Predictive analysis in healthcare systems. In A. J. Anand, K. Kalaiselvi & J. M. Chatterjee (Eds.), *Artificial Intelligence – Based System Models in Healthcare* (pp. 131–152). Wiley. https://doi.org/10.1002/9781394242528.ch6

Slonje, R., & Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, *49*(2), 147–154. https://doi.org/10.1111/j.1467-9450.2007.00611.x

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry*, *49*(4), 376–385. https://doi.org/10.1111/j.1469-7610.2007.01846.x

Smith, P. K., del Barrio, C., & Tokunaga, R. S. (2013). Definitions of bullying and cyberbullying: How useful are the terms? In S. Bauman, D. Cross & J. Walker (Eds.), *Principles of cyberbullying research: Definitions, measures, and methodology* (pp. 26–40). New York: Routledge/Taylor & Francis Group.

Spyropoulos, F. (2011). The manifestations of deviant behavior on the Internet – Reflections on the needs of modernization of Greek criminal legislation. In K. Siomos & G. Floros (Eds.), *Research, Prevention, Management of Risks in Internet Use*. Larissa: Hellenic Society for the Study of Internet Addiction Disorder. (In Greek).

Tavani, H. T. (2011). *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing* (3rd ed). Hoboken, NJ: Wiley.

Topcu-Uzer, C., & Tanrıkulu, İ. (2018). Technological solutions for cyberbullying. In M. Campbell & S. Bauman (Eds.), *Reducing cyberbullying in schools* (pp. 33–47). Academic Press. https://doi.org/10.1016/B978-0-12-811423-0.00003-1

Tsouramanis, Ch. (2005). *Digital Crime – The (Un)Safe Side of the Internet*. Athens: V. Katsaros Publications. (In Greek).

Vandebosch, H. (2019). Cyberbullying prevention, detection and intervention. In W. Heirman, M. Walrave & H. Vandebosch (Eds.), *Narratives in Research and Interventions on Cyberbullying among Young People* (pp. 29–44). Cham: Springer. https://doi.org/10.1007/978-3-030-04960-7_3

Vandebosch, H., & Van Cleemput, K. (2008). Defining cyberbullying: A qualitative research into the perceptions of youngsters. *CyberPsychology & Behavior*, *11*(4), 499–503. https://doi.org/10.1089/cpb.2007.0042

Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226852904.001.0001

Ybarra, M. L., & Mitchell, K. J. (2004a). Online aggressors/targets, aggressors, and targets: A comparison of associated youth characteristics. *Journal of Child Psychology and Psychiatry*, *45*, 1308–1316. https://doi.org/10.1111/j.1469-7610.2004.00328.x

Ybarra, M. L., & Mitchell, K. J. (2004b). Youth engaging in online harassment: Associations with caregiver-child relationships, Internet use, and personal characteristics. *Journal of Adolescence*, *27*, 319–336. https://doi.org/10.1016/j.adolescence.2004.03.007

Zannis, A. (2005). *Cybercrime*. Athens-Komotini: Sakkoulas Publications. (In Greek).

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

Author information

Fotios Spyropoulos – PostDoc, PhD, Associate Professor of Criminal Law & Criminology, Faculty of Law, Philips University; Senior Partner of Spyropoulos Law Firm
Address: 4-6 Lamias Street, 2001, P.O. Box 28008, Nicosia, Cyprus; Alexandras Avenue 81, 11474, Athens, Greece
E-mail: fspyropoulos@gmail.com
ORCID ID: https://orcid.org/0000-0001-5950-3583
Google Scholar ID: https://scholar.google.com/citations?user=iKQYLWoAAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – May 29, 2025
Date of approval – June 12, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:340.1.721:004.8
EDN: https://elibrary.ru/bvlgsu
DOI: https://doi.org/10.21202/jdtl.2025.19

Техноэтические и правовые аспекты киберзапугивания: структурный анализ дисбаланса сил в цифровой среде

Фотиос Спайропулос

Университет Филипс, Никосия, Кипр
Юридическая компания Spyropoulos Law Firm

Ключевые слова

алгоритмическая прозрачность,
искусственный интеллект,
киберзапугивание,
международное сотрудничество,
право,
технологическая грамотность,
цифровая безопасность,
цифровые платформы,
цифровые технологии,
этика

Аннотация

Цель: исследование направлено на концептуализацию понятия киберзапугивания с точки зрения права, техноэтики и анализ дисбаланса сил в цифровом пространстве как основополагающего фактора причинения вреда в Сети.

Методы: в работе применяется концептуально-аналитическая методология, базирующаяся на междисциплинарном анализе теоретических положений права, техноэтики, философии технологий и социальной психологии. Методологический инструментарий дополнен построением оригинальных концептуальных моделей на основе анализа структурных факторов цифрового пространства, разработкой схем причинно-следственных связей и созданием таксономии форм киберзапугивания. Особое внимание уделено компаративному анализу регулятивных подходов различных юрисдикций и выявлению пробелов в существующих правовых нормах.

Результаты: установлено, что киберзапугивание представляет собой сложный многоуровневый феномен, возникающий на пересечении архитектурных особенностей цифровых платформ, асимметрии технологических компетенций между участниками интеракций и системной фрагментированности законодательного регулирования. Выявлены критические пробелы в ключевых международных правовых инструментах, проявляющиеся в отсутствии унифицированных определений киберзапугивания, недостаточной проработке механизмов трансграничного сотрудничества и нерелевантном учете специфики цифровой среды. Проанализированы фундаментальные этические вопросы, связанные с автоматизированной модерацией контента на основе алгоритмов машинного обучения, проблематикой распределения ответственности между платформами, государственными регуляторами и индивидуальными пользователями, а также противоречиями между обеспечением безопасности и сохранением пользовательской автономии. Выделены четыре основных типа дисбаланса сил: технологический, информационный, социальный и институциональный, каждый из которых требует специфических стратегий преодоления.

Научная новизна: впервые предложен комплексный подход к анализу киберзапугивания как структурно обусловленного злоупотребления цифровой властью через призму техноэтики. Разработанные концептуальные модели представляют новые инструменты для понимания распределенной природы ответственности в цифровой экосистеме и формирования этически обоснованных стратегий профилактики. Введена концепция неправомерного использования информации как центрального механизма систематического злоупотребления властью в цифровой среде.

Практическая значимость: результаты исследования адресованы ученым-правоведам, государственным деятелям и разработчикам цифровых платформ, предлагая практические решения в области этического аудита алгоритмов, создания гибридных систем модерации с участием искусственного интеллекта и человека, формирования международных целевых групп и развития, основанных на правах человека принципов цифровой грамотности. Предложения автора направлены на создание более безопасной, подотчетной и инклюзивной цифровой среды для всех участников.

Для цитирования

Спайропулос, Ф. (2025). Техноэтические и правовые аспекты киберзапугивания: структурный анализ дисбаланса сил в цифровой среде. *Journal of Digital Technologies and Law*, *3*(3), 472–496. https://doi.org/10.21202/jdtl.2025.19

Список литературы

Bosworth, K., Espelage, D. L., & Simon, T. R. (1999). Factors associated with bullying behavior in middle school students. *The Journal of Early Adolescence*, *19*(3), 341–362. https://doi.org/10.1177/0272431699019003003

Chakraborty, S., Bhattacherjee, A., & Onuchowska, A. (2021). Cyberbullying: A review of the literature. SSRN. http://dx.doi.org/10.2139/ssrn.3799920

Campbell, M., & Bauman, S. (2018). Cyberbullying: Definition, consequences, prevalence. In M. Campbell & S. Bauman (Eds.), *Reducing Cyberbullying in Schools* (pp. 3–16). Academic Press. https://doi.org/10.1016/B978-0-12-811423-0.00001-8

Capurro, R. (2009). Digital ethics. *The Information Society*, *25*(3), 183–186. https://doi.org/10.1080/01972240902848902

Chen, C. W. Y. (2017). "Think before you type": The effectiveness of implementing an anti-cyberbullying project in an EFL classroom. *Pedagogies*, *13*(1), 1–18. https://doi.org/10.1080/1554480X.2017.1363046

Courakis, N. (2005). *Criminological horizons. Vol. II: Pragmatic approach and individual issues.* 2nd ed. Athens – Komotini: Ant. N. Sakkoula (In Greek).

Duncan, S. H. (2016). Cyberbullying and restorative justice. In R. Navarro, S. Yubero & E. Larrañaga (Eds.). *Cyberbullying Across the Globe* (pp. 239–257). Cham: Springer International Publishing.

Englander, E., Donnerstein, E., Kowalski, R., Lin, C. A., & Parti, K. (2017). Defining cyberbullying. *Pediatrics*, *140*(Supplement_2), S148–S151. https://doi.org/10.1542/peds.2016-1758u

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Furnell, S. M. (2006). *Computer Insecurity: Risking the System*. Oxford: Chandos Publishing.

Gibson, J. J. (2014). The theory of affordances (originally published 1979). In G. Bridge & S. Watson (Eds.). *The People, Place, and Space Reader* (pp. 56–60). London: Routledge.

Guardabassi, V., & Nicolini, P. (2024). Adolescence and cyberbullying: a restorative approach. In *INTED2024 Proceedings* (pp. 4562–4570). IATED. https://doi.org/10.21125/inted.2024.1183

Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. (2023). A review on deep-learning-based cyberbullying detection. *Future Internet*, *15*(5), 179. https://doi.org/10.3390/fi15050179

Hinduja, S., & Patchin, J. W. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice*, *4*(2), 148–169. https://doi.org/10.1177/1541204006286288

Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, CA: Sage Publications.

Ioannou, A., Blackburn, J., Stringhini, G., De Cristofaro, E., Kourtellis, N., & Sirivianos, M. (2018). From risk factors to detection and intervention: a practical proposal for future work on cyberbullying. *Behaviour & Information Technology*, *37*, 258–266. https://doi.org/10.1080/0144929X.2018.1432688

Johora, F. T., Khan, M. S. I., Kanon, E., Rony, M. A. T., Zubair, M., & Sarker, I. H. (2024). A data-driven predictive analysis on cyber security threats with key risk factors. *arXiv preprint arXiv:2404.00068*.

Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? Bullying experiences in cyberspace. *Journal of School Health*, *78*(9), 496–505. https://doi.org/10.1111/j.1746-1561.2008.00335.x

Katerelos, I., Tsekeris, C., Lavdas, M., & Demetriou, K. (2011). A psychosociological approach to Internet use and mass online role-playing games. In K. Siomos & G. Floros (Eds.), *Research, prevention, management of risks in Internet use*. Larissa: Hellenic Society for the Study of Internet Addiction Disorder. (In Greek).

Kim, S., Razi, A., Stringhini, G., Wisniewski, P. J. and De Choudhury, M. (2021). A human-centered systematic literature review of cyberbullying detection algorithms. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2). https://doi.org/10.1145/3476066

Kowalski, R. (2018). Cyberbullying. In J. P. Forgas, R. F. Baumeister & T. F. Denson (Eds.), *The Routledge International Handbook of Human Aggression* (pp. 131–142). London: Routledge. https://doi.org/10.4324/9781315618777-11

Langos, C. (2012). Cyberbullying: The challenge to define. *Cyberpsychology, Behavior, and Social Networking*, *15*(6), 285–289. https://doi.org/10.1089/cyber.2011.0588

Lazos, G. (2001). *Information Technology and Crime*. Athens: Nomiki Bibliothiki. (In Greek).

Li, Q. (2007). New bottle but old wine: A research of cyberbullying in schools. *Computers in Human Behavior*, *23*(4), 1777–1791. https://doi.org/10.1016/j.chb.2005.10.005

Luppicini, R. (2018). *The Changing Scope of Technoethics in Contemporary Society*. Hershey, PA: Information Science Reference. https://doi.org/10.4018/978-1-5225-5094-5

Meier, A., Reinecke, L., & Meltzer, C. E. (2016). "Facebocrastination"? Predictors of using Facebook* for procrastination and its effects on students' well-being. *Computers in Human Behavior*, *64*, 65–76. https://doi.org/10.1016/j.chb.2016.06.011

Menesini, E., Nocentini, A., & Palladino, B. E. (2012). Empowering students against bullying and cyberbullying: Evaluation of an Italian peer-led model. *International Journal of Conflict and Violence*, *6*(2), 313–320. https://doi.org/10.4119/ijcv-2922

Menesini, E., Nocentini, A., & Camodeca, M. (2013). Morality, values, traditional bullying, and cyberbullying in adolescence. *British Journal of Developmental Psychology*, *31*(1), 1–14. https://doi.org/10.1111/j.2044-835X.2011.02066.x

Millard, G. (2009). Stephen Harper and the politics of the bully. *Dalhousie Review*, *89*(3), 329–336.

Moor, J. H. (2005). Why we need better ethics for emerging technologies. *Ethics and Information Technology*, *7*(3), 111–119. https://doi.org/10.1007/s10676-006-0008-0

Mouzaki, D. (2010). International scientific conference on: "Dealing with cyberbullying from a legal perspective". *The Art of Crime*, *15*. (In Greek).

Nocentini, A., Calmaestra, J., Schultze-Krumbholz, A., Scheithauer, H., Ortega, R., & Menesini, E. (2010). Cyberbullying: Labels, behaviours and definition in three European countries. *Australian Journal of Guidance & Counselling*, *20*(2), 129–142. https://doi.org/10.1375/ajgc.20.2.129

Olweus, D. (1993). *Bullying at School: What We Know and What We Can Do*. Oxford, UK: Blackwell.

Ortega-Ruiz, R., Del Rey, R., & Casas, J. A. (2012). Knowing, building and living together on internet and social networks: the ConRed cyberbullying prevention program. *International Journal of Conflict and Violence*, *6*(2), 303–313. https://doi.org/10.4119/ijcv-2921

Othman, S. N., Alziboon, M. F., Dawood, M., Sachet, S. J., & Moroz, I. (2024). New rehabilitation against electronic crimes by young people. *Encuentros: Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico*, *22*, 363–385. https://doi.org/10.5281/zenodo.13732745

Raj, M., Singh, S., Solanki, K., & Selvanambi, R. (2022). An application to detect cyberbullying using machine learning and deep learning techniques. *SN Computer Science*, *3*(5), 401. https://doi.org/10.1007/s42979-022-01266-w

* Социальная сеть принадлежит Meta, признанной экстремистской организацией, ее деятельность запрещена на территории Российской Федерации.

Sathya, J., & Fernandez, F. M. H. (2024). Predictive analysis in healthcare systems. In A. J. Anand, K. Kalaiselvi & J. M. Chatterjee (Eds.), *Artificial Intelligence – Based System Models in Healthcare* (pp. 131–152). Wiley. https://doi.org/10.1002/9781394242528.ch6

Slonje, R., & Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, *49*(2), 147–154. https://doi.org/10.1111/j.1467-9450.2007.00611.x

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry*, *49*(4), 376–385. https://doi.org/10.1111/j.1469-7610.2007.01846.x

Smith, P. K., del Barrio, C., & Tokunaga, R. S. (2013). Definitions of bullying and cyberbullying: How useful are the terms? In S. Bauman, D. Cross & J. Walker (Eds.), *Principles of cyberbullying research: Definitions, measures, and methodology* (pp. 26–40). New York: Routledge/Taylor & Francis Group.

Spyropoulos, F. (2011). The manifestations of deviant behavior on the Internet – Reflections on the needs of modernization of Greek criminal legislation. In K. Siomos & G. Floros (Eds.), *Research, Prevention, Management of Risks in Internet Use*. Larissa: Hellenic Society for the Study of Internet Addiction Disorder. (In Greek).

Tavani, H. T. (2011). *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing* (3rd ed). Hoboken, NJ: Wiley.

Topcu-Uzer, C., & Tanrıkulu, İ. (2018). Technological solutions for cyberbullying. In M. Campbell & S. Bauman (Eds.), *Reducing cyberbullying in schools* (pp. 33–47). Academic Press. https://doi.org/10.1016/B978-0-12-811423-0.00003-1

Tsouramanis, Ch. (2005). *Digital Crime – The (Un)Safe Side of the Internet*. Athens: V. Katsaros Publications. (In Greek).

Vandebosch, H. (2019). Cyberbullying prevention, detection and intervention. In W. Heirman, M. Walrave & H. Vandebosch (Eds.), *Narratives in Research and Interventions on Cyberbullying among Young People* (pp. 29–44). Cham: Springer. https://doi.org/10.1007/978-3-030-04960-7_3

Vandebosch, H., & Van Cleemput, K. (2008). Defining cyberbullying: A qualitative research into the perceptions of youngsters. *CyberPsychology & Behavior*, *11*(4), 499–503. https://doi.org/10.1089/cpb.2007.0042

Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226852904.001.0001

Ybarra, M. L., & Mitchell, K. J. (2004a). Online aggressors/targets, aggressors, and targets: A comparison of associated youth characteristics. *Journal of Child Psychology and Psychiatry*, *45*, 1308–1316. https://doi.org/10.1111/j.1469-7610.2004.00328.x

Ybarra, M. L., & Mitchell, K. J. (2004b). Youth engaging in online harassment: Associations with caregiver-child relationships, Internet use, and personal characteristics. *Journal of Adolescence*, *27*, 319–336. https://doi.org/10.1016/j.adolescence.2004.03.007

Zannis, A. (2005). *Cybercrime*. Athens-Komotini: Sakkoulas Publications. (In Greek).

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

Сведения об авторе

Спайропулос Фотиос – научный сотрудник, PhD, доцент в области уголовного права и криминалистики, факультет права, Университет Филипс; старший партнер юридической компании Spyropoulos Law Firm
Адрес: Кипр, 28008, г. Никосия, ул. Ламиас, д. 4-6; Греция, 11474, г. Афины, Александрас авеню, д. 81
E-mail: fspyropoulos@gmail.com
ORCID ID: https://orcid.org/0000-0001-5950-3583
Google Scholar ID: https://scholar.google.com/citations?user=iKQYLWoAAAAJ

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс
Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 29 мая 2025 г.
Дата одобрения после рецензирования – 12 июня 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.

Research article
UDC 34:004:347:004.4
EDN: https://elibrary.ru/jqhnur
DOI: https://doi.org/10.21202/jdtl.2025.20

Electronic Evidence as an Independent Legal Phenomenon: Theoretical Framework and European Practice

Gergana Varbanova
Nikola Vaptsarov Naval Academy, Varna, Bulgaria

Keywords

authenticity,
digital information,
digital technologies,
electronic evidence,
European legislation,
evidence,
judicial procedure,
law,
material evidence,
procedure

Abstract

Objective: to develop a new theoretical framework that challenges the traditional classification of electronic evidence as a subtype of material evidence and suggests considering it as a qualitatively new legal phenomenon with its own independent legal nature in the context of applicable European legislation.

Methods: the work uses the doctrinal method for the legal analysis of applicable European legislation, including Regulation (EU) 2023/1543 and Regulation (EU) 910/2014 (eIDAS), as well as their direct application in the national legal systems of the European Union member states. A comparative legal approach was used to identify differences between theoretical views and case law. A technological analysis of digital information was performed; specific examples were explained to illustrate the problems associated with the collection and use of electronic evidence within the European legislation framework.

Results: the author proposes a new doctrinal understanding of electronic evidence as an independent category that differs from traditional material evidence in its digital nature and specific characteristics. The introduction of European regulations requires rethinking the legal nature of electronic evidence as a qualitatively different legal phenomenon. It was established that treating electronic evidence as material one creates a risk of legal uncertainty, while the lack of appropriate legal regulation hinders effective law enforcement.

Scientific novelty: for the first time, the research proposes to overcome the established paradigm and identify electronic evidence as an independent

legal category in the system of evidence types. The article substantiates the unique digital nature of electronic evidence and the need to create an independent legal framework in various national legislations. It is proposed to improve scientific terminology using the term “electronic evidence”, which corresponds to the legal definitions in the legislation under study, instead of the outdated term “digital evidence”.

Practical significance: the work contains specific practical recommendations for the use of electronic evidence in the procedures for its identification, storage, presentation and analysis in various court proceedings in accordance with applicable supranational legislation. The research helps to overcome outdated ideas about the legal nature of electronic evidence and their incorrect identification with material evidence. This is important for effective law enforcement in the European Union member states.

For citation

Varbanova, G. (2025). Electronic Evidence as an Independent Legal Phenomenon: Theoretical Framework and European Practice. *Journal of Digital Technologies and Law, 3*(3), 497–511. https://doi.org/10.21202/jdtl.2025.20

Contents

Introduction

Electronic evidence has a special, different nature from other evidence that exists in the analog world and that certifies facts that occurred in the past, but are relevant to the present. The peculiarity of electronic evidence is due to the fact that it can remain unchanged over a long period of time, but can be dynamic, as its content can be different from the moment when the fact it certifies existed. It is difficult to trace, it can often be changed in order to conceal information or to deliberately alter it. Electronic evidence comprises digital information encoded as binary data (a sequence of ones and zeros), and it challenges legal theory and practice, intertwining legal knowledge with the technical features of this new legal phenomenon – electronic evidence. Digital information cannot

be considered as a material object, it cannot be perceived directly, it is not an object from the physical world, and this poses serious challenges to legal theory and practice.

This study proposes a new theoretical framework that challenges the traditional classification of electronic evidence as a subset of physical evidence. The author proposes a change in the paradigmatic understanding of the legal nature of electronic evidence and considers it as a new legal phenomenon that has its own independent legal nature and supranational regulation. The study focuses only on the applicable European legislation, including Regulation (EU) 2023/1543, Regulation (EU) 910/2014 (eIDAS), and the challenges that this new theory of the legal nature of electronic evidence poses.

The study uses the doctrinal method for legal analysis of the applicable European regulations and their direct application in the national legal systems of the Member States. A comparative approach is applied to clarify the differences between theoretical views and case law arising in the application of supranational legal norms. The analysis examines the technological aspects related to the nature of digital information and presents specific examples that illustrate the challenges in collecting and using electronic evidence in the context of European regulations.

This paper presents an effort to offer ways to overcome the old paradigm and to separate electronic evidence as an independent legal category in the system of types of evidence, as well as to distinguish it from physical evidence. The study presents arguments for the unique digital nature of electronic evidence and the need to create an independent legal framework in various national legislations. It is proposed to refine the scientific terminology by using the term «electronic evidence», which corresponds to the legal definitions given in the regulations under consideration, and to overcome the use of the outdated term «digital evidence».

1. The classical theory of physical evidence

The theory defines physical evidence as material objects that can reproduce facts relevant to the case or allow for the drawing of evidentiary conclusions about these facts. They refer to events that occurred in the past, but which have significance in the present and are related to the process of proving. For example, Mubarik (2019) considers as physical evidence any object, material, fingerprints or various substances (including bodily) collected from the crime scene that are relevant to the investigation and can contribute to clarifying the facts of the case. In order for an evidence to be qualified as physical, it is necessary that it has a material nature and can be perceived through one of the human senses . Some theorists argue that the physical medium carrying digital information. (e. g. disk, USB, hard drive) can be considered as physical evidence (Pastukhov, 2015; Dmitrieva & Pastukhov, 2023).

This traditional theory has shortcomings and is subject to serious criticism, because electronic evidence, considered as digital information, does not always have a material nature and can exist entirely in a digital environment, without being objectified on a physical medium – for example, records on a cloud server. Electronic evidence often does not exist in analog form and is not part of the material world in the classical sense of the term.

2. European regulatory framework applicable to electronic evidence

In an effort to define and unify the understanding of the specific legal nature of electronic evidence, the European Union adopted Regulation (EU) 2023/1543, which has direct application and provides a legal definition of the concept of "electronic evidence". Such a legislative decision is not accidental, but aims to overcome the different interpretations of the concepts in the national legislation of the Member States and to have a unified understanding on the legal nature of electronic evidence.

The Regulation defines "electronic evidence" as subscriber data, traffic data or content data stored by a service provider or on behalf of a service provider in electronic form. The legislative approach is neutral and does not provide an exhaustive list of all types of digital information that could be qualified as electronic evidence. The Regulation specifically focuses only on electronic evidence – digital information that is stored or transmitted by service providers in the context of their provision of digital services and for the purposes of criminal proceedings. Files or other digital data stored locally by users do not fall within the scope of Regulation (EU) 2023/1543. Such data may fall under the definition of "electronic documents" as per Regulation (EU) 910/2014 (eIDAS). when they contain information that is legally relevant for various civil, commercial, administrative or other public relations.

Regulation (EU) No 910/2014 (eIDAS) defines an "electronic document" as any content stored in electronic form, in particular in the form of text, sound, visual or audiovisual recording. In the context of eIDAS, an electronic document is a carrier of information in digital form which may be relevant to civil, administrative, commercial or other public relations. It is sufficient that the electronic document contains information which is legally relevant, regardless of whether the document was created accidentally or intentionally for the purposes of a particular legal relationship. Article 46 of Regulation (EU) No 910/2014 provides that an electronic document has the same evidentiary value as other evidence. The cited norm imperatively states that the evidentiary value of an electronic document cannot be denied solely because it is in electronic form, which, in turn, ensures certainty in cross-border disputes, including arbitration proceedings (Ferreira & Gromova, 2024). The Regulation practically obliges Member States

to recognize the validity of electronic documents as admissible evidence in various legal proceedings (Nekrošius, 2021; Daniel & Daniel, 2012)[1].

In the context of the study, the concepts of "electronic evidence" within the meaning of Regulation (EU) 2023/1543 and "electronic document" under Regulation (EU) 910/2014 should be compared. Electronic evidence under Regulation (EU) 2023/1543 covers a wide range of digital information – subscriber data, traffic data and content data, which may include structured or unstructured data, metadata or content of communications – and is distinct from electronic documents. The scope of Regulation (EU) 2023/1543 is limited to specific cases in criminal proceedings – for the preservation and provision of electronic evidence, and only in relation to digital information stored by or on behalf of service providers.

On the other hand, the concept of "electronic document" within the meaning of Regulation (EU) 910/2014 (eIDAS) has a narrower technological application, as it essentially refers to digital content perceived as a "document", but its legal applicability is significantly broader. The "electronic document" is not limited to the field of criminal proceedings, nor to data stored by specific service providers and is applicable across all branches of law: civil, commercial, administrative, and criminal. (Kirkov, 2022; Berube et al., 2025; Shin et al., 2025). Therefore, it can be concluded that electronic evidence under Regulation 2023/1543 represents a specialized type of digital information intended for the purposes of criminal proceedings, and has a narrower scope of regulation. In contrast, electronic documents under Regulation 910/2014 have a broader scope and significance, as their admissibility as evidence does not depend on the specific context of criminal proceedings, nor on the place of storage or the source of the data, but on compliance with the principles established by EU law for electronic identification, authentication and trust services, including the types of electronic signatures used in electronic documents.

It is noteworthy that the definition of "content data" under Regulation (EU) 2023/1543 and the definition of "electronic documents" under eIDAS are similar. The difference lies in the context and objectives, as "content data" under Regulation 2023/1543 is oriented towards digital information that is transmitted and stored by digital service providers and that should be preserved, respectively provided for the needs of criminal justice, while the definition of "electronic document" under eIDAS is related to the recognition of the legal value and use of electronic documents in all areas of public life.

The definition of electronic evidence within the meaning of Regulation (EU) 2023/1543 is not exhaustive, but only outlines some of the types of digital information that can be treated as electronic evidence. The approach is identical in the definition of the concept of "electronic document" under Regulation (EU) 910/2014 – any content stored in electronic

[1] Varbanova, G. (2020). Legal regime of electronic documents. Varna: Dangrafik Publishing.

form, in particular text, sound, visual or audiovisual recording. In both cases, the list is not exhaustive, taking into account that technologies are developing. When regulating public relations in the field of information technology, the legislative approach must be flexible, given the dynamics of technological development. Technological development implies the emergence of new types of electronic evidence and electronic documents, which cannot be excluded from the scope of Regulation (EU) 2023/1543 or eIDAS simply because they are not explicitly defined as electronic evidence or electronic documents, respectively.

3. Definition and legal nature of electronic evidence

Electronic evidence is a key tool in the evidentiary process, which requires a special approach to its collection, analysis and legal assessment. This approach must take into account the intensive development of technologies, as well as the specific features of electronic evidence in the context of the rule-making process (Begishev et al., 2020).

Until the adoption of Regulation (EU) 2023/1543, the theory defined electronic evidence as any information stored or transmitted in digital form that can be used as evidence. Traditionally, some authors assume that electronic evidence is physical evidence, since the information it contains is objectified on a specific material medium. According to this approach, it is the physical characteristics of the medium that determine the evidentiary value, and not the information itself (electronic evidence) recorded on it. Although initially plausible, this concept is flawed, because it does not take into account and does not reflect the specific nature of electronic evidence – as digital information, which has its own unique characteristics (Wu et al., 2025)[2].

In more recent research, there is also a theory that electronic documents are a special category of quasi-material evidence. According to (Bufetova, 2023; Guo, 2022), an electronic document as material evidence is a document that exists in electronic form, contains information relevant to the case, and is recorded on an electronic medium that allows the reproduction and use of this information in the process of proving. The very definition given by the author shows that an electronic document is considered a specific type – quasi-material evidence, in which the information contained in the material medium is important, and not its material characteristics.

This confusion arises from a misunderstanding of how digital information is generated, modified, stored and deleted. Digital information can be stored on various media – a hard drive, a USB device, a cloud, a server – or transmitted via electronic channels, but the information is not identical to the medium itself. A file can be stored on a computer or other technical medium, copied, transferred or deleted, but the computer

[2] Varbanova, G. (2024). The significance of electronic evidence in the context of cybersecurity and national security. Print Master Publishing.

or medium on which it is recorded does not constitute physical evidence, especially since the information may be recorded on a cloud server and accessed via a computer or other device.

From the definitions of electronic evidence and electronic documents it is evident that they do not have a material (tangible) nature, but represent a digital record, therefore they can be defined as a variety of intangible evidence (Vuchkov, 2023; Horsman, 2021). In order to be able to define electronic evidence as a new legal phenomenon, it is necessary to clarify their digital nature and the way in which digital information is created, transferred, stored, recorded and deleted.

Digital information is data in binary code – a series of zeros and ones processed by information systems (computer, smart device, etc.). Thus, text, images, sound or video recorded on a technical medium represent digital information in binary code that can be perceived by human senses through the use of generally accepted standards for converting and reproducing information, through which the zeros and ones are visualized as text, images or audio files.

To be admissible in court, the collection of electronic evidence must ensure the full integrity and identity of the information data, so as to guarantee its authenticity, the integrity of the content and the immutability of the data. Only when these requirements are met can electronic evidence be accepted as valid and reliable evidence in court proceedings.

4. Authenticity, integrity, and evidentiary value of electronic documents

The modern world is characterized by a continuous increase in legal relationships that arise, develop and terminate in an electronic environment. Every day, numerous contracts are concluded entirely electronically, consumers use online services, make electronic payments, and even legal proceedings are now carried out via videoconferencing and using tools of the electronic world. Almost all spheres of public and economic life are closely related to the use of electronic means of communication, processing and storage of information.

Cyberspace has become a commercial zone that transcends physical boundaries, but at the same time represents a center of attraction for committing numerous new and previously unknown computer crimes.

Electronic evidence – whether it is electronic documents, communication records, log files, metadata or other forms of digital information – is an integral part of legal relations in an electronic environment and is of essential importance for revealing crimes committed in cyberspace. Proving the facts, whether it is a question of civil or commercial legal relations, or of crimes committed in cyberspace, requires guaranteeing the authenticity, integrity and immutability of electronic evidence. This is achieved by implementing appropriate technical and organizational measures.

A key challenge in the presentation, analysis and admission of electronic evidence in legal proceedings – whether civil, criminal or administrative – is the authenticity, integrity and probative value of electronic evidence and electronic documents. In today's digital world, electronic evidence is becoming the most commonly used evidence. In order to be accepted as valid and reliable means of evidence, electronic evidence and electronic documents must satisfy several criteria, including the assurance of digital data integrity, in order to guarantee their authenticity, integrity and immutability.

An electronic document is considered authentic when its author, place and time of creation are established, and its content actually originates from the stated author and has not been altered since its creation (Surovtseva, 2020). In this sense, the authenticity and immutability of the electronic document can be ensured through the use of certification services under Regulation (EU) 910/2014 (eIDAS) – for example, a qualified electronic signature, a qualified electronic seal or a qualified electronic time stamp, which certify both the identity of the author and the time of creation and the integrity of the electronic document. The integrity of the electronic document or digital information can also be guaranteed through the use of blockchain technology and smart contracts, which provide the ability to store data in a decentralized, immutable and transparent environment (Miao et al., 2021). The blockchain ledger and its records ensure that the content of the electronic document has not been altered since its initial creation. The technology provides secure storage, protection against subsequent modification or deletion of the electronic document, as well as chronological verification of the records in the decentralized ledger, which guarantees the authenticity and traceability of electronic evidence (Al-E'mari et al., 2024; Stoykova, 2023).

Artificial intelligence (AI) systems can also be used in the evidence process, and in particular in the authentication phase of electronic documents, through automated content analysis, anomaly detection, version comparison and risk assessment, which enhances the security of handling electronic evidence. The integration of AI systems, blockchain technologies and qualified certification services creates a multi-layered mechanism for protecting the integrity and authenticity of electronic documents and electronic evidence, especially in the context of litigation and e-justice.

Conclusions

This study proposes a change in the traditional paradigm of understanding the legal nature of electronic evidence in the context of European legislation. Regulation (EU) 2023/1543 and Regulation (EU) 910/2014 (eIDAS) are analyzed, and the method of comparative legal analysis justifies the conclusion that electronic evidence cannot be equated with the traditional category of physical evidence, since it has a special, digital nature and unique technical characteristics. Electronic evidence is a new legal phenomenon that

must receive an independent legal regulation regarding the procedural actions related to its collection, storage, analysis and acceptance as a suitable means of evidence in the legal process. Considering electronic evidence as a type of physical evidence is not only incorrect, but also creates a real risk of legal uncertainty and difficulties in effective law enforcement in individual Member States. The study contributes to the development of the theoretical understanding of electronic evidence, consistent with applicable European law, while outlining opportunities for future research on the integration of emerging technologies such as blockchain and artificial intelligence in the process of proving and verifying the authenticity of electronic evidence.

The study proposes to overcome the old paradigm and to separate electronic evidence as an independent legal category in the system of types of evidence, as well as to distinguish it from physical evidence. The study presents arguments for the unique digital nature of electronic evidence and the need to create an independent legal framework in various national legislations. It is proposed to refine the scientific terminology by using the term «electronic evidence», which corresponds to the legal definitions given in the regulations under consideration, and to overcome the use of the outdated term «digital evidence».

The study provides specific guidelines of practical importance in the process of using electronic evidence – for the identification, storage, presentation and analysis of electronic evidence in various legal proceedings, in accordance with the applicable supranational law. At the same time, this study proposes to overcome outdated concepts regarding the legal nature of electronic evidence and its incorrect identification with physical evidence.

References

Al-E'mari, S., Sanjalawe, Y., Makhadmeh, S., & Alqudah, H. (2024). Digital forensics meets blockchain: Enhancing evidence authenticity and traceability. *2024 2nd International conference on cyber resilience (ICCR)* (pp. 01–06). IEEE. https://doi.org/10.1109/ICCR61006.2024.10532961

Begishev, I. R., Khisamova, Z. I., & Nikitin, S. G. (2020). The organization of hacking community: Criminological and criminal law aspects. *Russian Journal of Criminology*, *14*(1), 96–105. (In Russ.). https://doi.org/10.17150/2500-4255.2020.14(1).96-105

Bérubé, M., Beaulieu, L., Allard, S., & Denault, V. (2025). From digital trace to evidence: Challenges and insights from a trial case study. *Science & Justice*, *65*(5), 101306. https://doi.org/10.1016/j.scijus.2025.101306

Bufetova, M. Sh. (2023). The electronic document as a type of physical evidence in Russian criminal proceedings. *Siberian Legal Readings*, *3*, 47–55. https://doi.org/10.17150/2411-6122.2023.3.47-55

Daniel, Larry. E., & Daniel, Lars. E. (2012). *Discovery of digital evidence in civil cases. In Digital Forensics for Legal Professionals* (Ch. 16, pp. 113–121). Elsevier eBooks. https://doi.org/10.1016/b978-1-59749-643-8.00016-x

Dmitrieva, A. A., & Pastukhov, P. S. (2023). Concept of Electronic Evidence in Criminal Legal Procedure. *Journal of Digital Technologies and Law*, *1*(1), 270–295. https://doi.org/10.21202/jdtl.2023.11

Ferreira, D., & Gromova, E. (2024). Digital evidence in disputes involving states. *AJIL Unbound*, *118*, 51–56. https://doi.org/10.1017/aju.2024.4

Guo, Z. (2022). Regulating the use of electronic evidence in Chinese courts: Legislative efforts, academic debates and practical applications. *Computer Law & Security Review*, *48*, 105774. https://doi.org/10.1016/j.clsr.2022.105774

Horsman, G. (2021). Digital evidence and the crime scene. *Science & Justice*, *61*(6), 761–770. https://doi.org/10.1016/j.scijus.2021.10.003

Kirkov, A. (2022). Forensic activity experts in Bulgaria. Organization and problems. In *Collection of materials international scientific and practical seminar on the topic: "Digitalization of activities Courts: Current and Prospective Tasks"* (AI No. 276, 15.06.2015, pp. 5–9).

Miao, Z., Ye, C., Yang, P., Chen, Y., & Chen, Y. (2021). Blockchain-based Electronic Evidence Storage and Efficiency Optimization. In *2021 International Conference on Artificial Intelligence and Blockchain Technology (AIBT)* (pp. 109–113). IEEE. https://doi.org/10.1109/AIBT53261.2021.00025

Mubarik, N. (2019). Importance and role of physical evidence in forensic science. *Academic Journal of Forensic Sciences*, *2*(1), 18–22.

Nekrošius, V. (2021). Use of information technologies in Lithuanian civil procedure. *Procedia Computer Science*, *192*, 2662–2667. https://doi.org/10.1016/j.procs.2021.09.036

Pastukhov, P. S. (2015). Elektronnoe important evidence in a criminal case sudoproizvodstve. *Vestnik Tomsk state-owned University*, *396*, 149–153.

Shin, D., Ha, J., & Euom, I. (2025). Data-Driven Digital evidence analysis for the forensic investigation of the electric vehicle charging infrastructure. *Computer Modeling in Engineering & Sciences*, *143*(3), 3795–3838. https://doi.org/10.32604/cmes.2025.066727

Stoykova, R. (2023). The right to a fair trial as a conceptual framework for digital evidence rules in criminal investigations. *Computer Law & Security Review*, *49*, 105801. https://doi.org/10.1016/j.clsr.2023.105801

Surovtseva, N. (2020). Authenticity and identity of electronic records. *Herald of an archivist*, *2*, 467–477. https://doi.org/10.28995/2073-0101-2020-2-467-477

Vuchkov, V. (2023). Electronic evidence in criminal cases and its sources – necessary considerations. Law *Journal of New Bulgarian University*, *19*(2), 12–19. https://doi.org/10.33919/ljnbu.23.2.1

Wu, D., Zuo, X., & Guo, Y. (2025). The application of electronic evidence in forensic imaging analysis. *Journal of Radiation Research and Applied Sciences*, *18*(3), 101708. https://doi.org/10.1016/j.jrras.2025.101708

Author information

Gergana Varbanova – PhD, Assistant Professor, Department of National Security and Information Technology Law, Nikola Vaptsarov Naval Academy
Address: 73 Vasil Drumev Street, Varna, Bulgaria
E-mail: g.varbanova@naval-acad.bg
ORCID ID: https://orcid.org/0000-0001-8122-4353
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=60021317100
WoS Researcher ID: https://www.webofscience.com/wos/author/record/HKP-1334-2023
Google Scholar ID: https://scholar.google.com/citations?hl=en&user=02-0uFYAAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – June 99, 2025
Date of approval – June 28, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:347:004.4
EDN: https://elibrary.ru/jqhnur
DOI: https://doi.org/10.21202/jdtl.2025.20

Электронные доказательства как самостоятельный правовой феномен: теоретические основы и европейская практика

Гергана Варбанова

Высшее военно-морское училище имени Николы Вапцарова, Варна, Болгария

Ключевые слова

аутентичность,
вещественные доказательства,
доказательства,
европейское законодательство,
право,
процессуальные действия,
судебное разбирательство,
цифровая информация,
цифровые технологии,
электронные доказательства

Аннотация

Цель: исследование направлено на разработку новой теоретической основы, которая бросает вызов традиционной классификации электронных доказательств как подвида вещественных доказательств и предлагает рассматривать их как качественно новый правовой феномен с собственной независимой правовой природой в контексте применимого европейского законодательства.

Методы: в работе применяется доктринальный метод для юридического анализа применимых европейских нормативных актов, включая Регламент (ЕС) 2023/1543 и Регламент (ЕС) 910/2014 (eIDAS), и их непосредственного применения в национальных правовых системах государств – членов Европейского союза. Для выявления различий между теоретическими взглядами и прецедентным правом используется сравнительно-правовой подход. Проводится технологический анализ цифровой информации и поясняются конкретные примеры, иллюстрирующие проблемы, связанные со сбором и использованием электронных доказательств в рамках европейского законодательства.

Результаты: автор предлагает новое доктринальное понимание электронных доказательств как самостоятельной категории доказательств, отличающейся от традиционных вещественных доказательств цифровой природой и специфическими характеристиками. Внедрение европейских регламентов требует переосмысления правовой природы электронных доказательств как качественно отличного правового явления. Установлено, что отношение к электронным доказательствам как к вещественным создает риск правовой неопределенности, а отсутствие соответствующего правового регулирования препятствует эффективному правоприменению.

Научная новизна: в исследовании впервые предлагается преодолеть устоявшуюся парадигму и выделить электронные доказательства как самостоятельную правовую категорию в системе видов доказательств. Обосновывается уникальная цифровая природа электронных

доказательств и необходимость создания независимой правовой базы в различных национальных законодательствах. Предложено совершенствование научной терминологии с использованием термина «электронные доказательства», соответствующего юридическим определениям в рассматриваемых нормативных правовых актах, вместо устаревшего термина «цифровые доказательства».

Практическая значимость: работа содержит конкретные рекомендации практического характера для использования электронных доказательств в процедурах их идентификации, хранения, представления и анализа в различных судебных разбирательствах в соответствии с применимым наднациональным законодательством. Исследование способствует преодолению устаревших представлений о правовой природе электронных доказательств и их неверного отождествления с вещественными доказательствами, что имеет важное значение для эффективного правоприменения в государствах – членах Европейского союза.

Для цитирования

Варбанова, Г. (2025). Электронные доказательства как самостоятельный правовой феномен: теоретические основы и европейская практика. *Journal of Digital Technologies and Law*, *3*(3), 497–511. https://doi.org/10.21202/jdtl.2025.20

Список литературы

Al-E'mari, S., Sanjalawe, Y., Makhadmeh, S., & Alqudah, H. (2024). Digital forensics meets blockchain: Enhancing evidence authenticity and traceability. *2024 2nd International conference on cyber resilience (ICCR)* (pp. 01–06). IEEE. https://doi.org/10.1109/ICCR61006.2024.10532961

Begishev, I. R., Khisamova, Z. I., & Nikitin, S. G. (2020). The organization of hacking community: Criminological and criminal law aspects. *Russian Journal of Criminology*, *14*(1), 96–105. https://doi.org/10.17150/2500-4255.2020.14(1).96-105

Bérubé, M., Beaulieu, L., Allard, S., & Denault, V. (2025). From digital trace to evidence: Challenges and insights from a trial case study. *Science & Justice*, *65*(5), 101306. https://doi.org/10.1016/j.scijus.2025.101306

Bufetova, M. Sh. (2023). The electronic document as a type of physical evidence in Russian criminal proceedings. *Siberian Legal Readings*, *3*, 47–55. https://doi.org/10.17150/2411-6122.2023.3.47-55

Daniel, Larry. E., & Daniel, Lars. E. (2012). *Discovery of digital evidence in civil cases. In Digital Forensics for Legal Professionals* (Ch. 16, pp. 113–121). Elsevier eBooks. https://doi.org/10.1016/b978-1-59749-643-8.00016-x

Dmitrieva, A. A., & Pastukhov, P. S. (2023). Concept of Electronic Evidence in Criminal Legal Procedure. *Journal of Digital Technologies and Law*, *1*(1), 270–295. https://doi.org/10.21202/jdtl.2023.11

Ferreira, D., & Gromova, E. (2024). Digital evidence in disputes involving states. *AJIL Unbound*, *118*, 51–56. https://doi.org/10.1017/aju.2024.4

Guo, Z. (2022). Regulating the use of electronic evidence in Chinese courts: Legislative efforts, academic debates and practical applications. *Computer Law & Security Review*, *48*, 105774. https://doi.org/10.1016/j.clsr.2022.105774

Horsman, G. (2021). Digital evidence and the crime scene. *Science & Justice*, *61*(6), 761–770. https://doi.org/10.1016/j.scijus.2021.10.003

Kirkov, A. (2022). Forensic activity experts in Bulgaria. Organization and problems. In *Collection of materials international scientific and practical seminar on the topic: "Digitalization of activities Courts: Current and Prospective Tasks"* (AI No. 276, 15.06.2015, pp. 5–9).

Miao, Z., Ye, C., Yang, P., Chen, Y., & Chen, Y. (2021). Blockchain-based Electronic Evidence Storage and Efficiency Optimization. In *2021 International Conference on Artificial Intelligence and Blockchain Technology (AIBT)* (pp. 109–113). IEEE. https://doi.org/10.1109/AIBT53261.2021.00025

Mubarik, N. (2019). Importance and role of physical evidence in forensic science. *Academic Journal of Forensic Sciences*, *2*(1), 18–22.

Nekrošius, V. (2021). Use of information technologies in Lithuanian civil procedure. *Procedia Computer Science*, *192*, 2662–2667. https://doi.org/10.1016/j.procs.2021.09.036

Pastukhov, P. S. (2015). Elektronnoe important evidence in a criminal case sudoproizvodstve. *Vestnik Tomsk state-owned University*, *396*, 149–153.

Shin, D., Ha, J., & Euom, I. (2025). Data-Driven Digital evidence analysis for the forensic investigation of the electric vehicle charging infrastructure. *Computer Modeling in Engineering & Sciences*, *143*(3), 3795–3838. https://doi.org/10.32604/cmes.2025.066727

Stoykova, R. (2023). The right to a fair trial as a conceptual framework for digital evidence rules in criminal investigations. *Computer Law & Security Review*, *49*, 105801. https://doi.org/10.1016/j.clsr.2023.105801

Surovtseva, N. (2020). Authenticity and identity of electronic records. *Herald of an archivist*, *2*, 467–477. https://doi.org/10.28995/2073-0101-2020-2-467-477

Vuchkov, V. (2023). Electronic evidence in criminal cases and its sources – necessary considerations. Law *Journal of New Bulgarian University*, *19*(2), 12–19. https://doi.org/10.33919/ljnbu.23.2.1

Wu, D., Zuo, X., & Guo, Y. (2025). The application of electronic evidence in forensic imaging analysis. *Journal of Radiation Research and Applied Sciences*, *18*(3), 101708. https://doi.org/10.1016/j.jrras.2025.101708

Информация об авторе

Варбанова Гергана– PhD, ассистент кафедры права в области национальной безопасности и информационных технологий, Высшее военно-морское училище имени Николы Вапцарова
Адрес: Болгария, г. Варна, ул. Василя Друмева, д. 73
E-mail: g.varbanova@naval-acad.bg
ORCID ID: https://orcid.org/0000-0001-8122-4353
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=60021317100
WoS Researcher ID: https://www.webofscience.com/wos/author/record/HKP-1334-2023
Google Scholar ID: https://scholar.google.com/citations?hl=en&user=02-0uFYAAAAJ

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.91 / Государство и право отдельных стран
Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 19 июня 2025 г.
Дата одобрения после рецензирования – 28 июня 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.

Research article
UDC 34:004:343.721:004.8
EDN: https://elibrary.ru/tnqlxy
DOI: https://doi.org/10.21202/jdtl.2025.21

Impact of Cybercrime Related Offences on E-commerce in Cameroon: Social-Legal and Economic Consequences

Anoumbuandem Benvolio Lekunze
University of Buea, Buea, Cameroon

Keywords

Cameroon,
cybercrime,
digital technologies,
e-commerce,
fraud,
justice,
law,
scamming,
security,
transactions

Abstract

Objective: to examine the impact of cybercrimes on e-commerce related transactions in Cameroon and evaluate the effectiveness of the legal provisions in force that counteract cyberthreats.

Methods: The research is based on the utilitarian, transaction cost and the rational choice theories. It adopts the qualitative research methodology with the use of the doctrinal method. The author conducted a comprehensive analysis of Cameroon's legal acts in the field of cybersecurity and e-commerce. A survey was carried out between January to April 2025 at Molyko in Buea where 250 sample responses were obtained. Judicial precedents and statistics of the Cameroon Ministry of Posts and Telecommunications were investigated.

Results: It was found that cybercrimes have caused loss of trust and confidence in e-commerce transactions within Cameroon and a declining rate at which people are willing to carry out e-commerce transactions in Cameroon. More than 60% of young persons between the ages of 16 to 35 years in some major Cameroonian cities are either involved in e-commerce related cybercrimes or suffered from them. It was also observed that there is an increase in the rate at which female persons are involved in e-commerce related cybercrimes. The main types of cybercrimes were identified: scamming, phishing, and bank card skimming.

Scientific novelty: it consists in a comprehensive interdisciplinary analysis of the impact of cybercrime on e-commerce in the context of the developing African economy. For the first time, an empirical study of the scale of cybercrime in a specific region of Cameroon was conducted, including

a quantitative assessment of youth involvement in illegal activities. The author has developed a theoretical model that combines the utilitarianism, transaction costs, and rational choice concepts to explain the motivation of cybercriminals. Specific socio-legal factors contributing to the growth of cybercrime in the context of the socio-political crisis were identified.

Practical significance: The study results are of great practical significance for improving the legal, technological, social and economic mechanisms for countering cybercrime in Cameroon. The proposed recommendations include reforming procedural legislation, expanding the powers of specialized agencies, introducing a system of home addresses and social security numbers, raising the minimum wage, and integrating courses on cybersecurity into educational programs. The data obtained can be used by government agencies, the judicial system, educational institutions and international organizations to develop effective strategies to combat cybercrime and develop a secure digital economy.

For citation

Lekunze, A. B. (2025). Impact of Cybercrime Related Offences on E-commerce in Cameroon: Social-Legal and Economic Consequences. *Journal of Digital Technologies and Law, 3*(3), 512–536. https://doi.org/10.21202/jdtl.2025.21

Contents

Introduction

Crimes are frequent occurrences in every human society, old as the creation of humans, and operate as a core concept in modern society (Chris et al., 2005). The introduction of the internet to the public on January 1, 1983[1] has facilitated the commission of some crimes through online communication portals that were not initially foreseen by the legislator. The first recorded history of cybercrime was in 1834 when the French telegraph system was hacked, and access gained to financial markets through stolen data[2]. The act of committing cybercrime involves different elements, tools and processes as opposed to conventional crimes that may not raise many issues of jurisdiction and proof. This is because cyberspace is vast and unlimited to a geographical region or country (Yokotani & Takano, 2022; 2021).

Cybercrime can be defined as a wide range of criminal activities that are carried out using digital devices and networks (Alhadidi et al., 2024; Arroyabe et al., 2024; Edwards & Hollely, 2023; Gupta et al., 2025; Higgs & Flowerday, 2025a)[3]. It is also a collection of criminal activities that include offences against computers and computer systems (Payne, 2020). A computer can be the object of commission or a target. Cybercrime can also be any criminal activity that uses a computer as either an instrumentality, or a means for perpetuating further crimes. It can take the form of cyber theft or other related illegal activities targeted on humans and property (Garner, 1999). This article focuses on the impact of cybercrimes in electronic commerce (e-commerce) transactions in Cameroon because available literature on the impact of cybercrimes n oe-commerce in Cameroon is limited. Examples of cybercrimes include but not limited to, hacking, cyber stalking, defamation, email bombing, data diddling, salami attacks, denial of service attack, virus and worm attacks, internet time thefts etc. E-commerce transactions as the name signify operates in cyberspace by peer-to-peer electronic data interchange (EDI) (Garner, 1999) and not always void of crimes.

---

1 January 1, 1983 is considered the official birthday of the Internet. Prior to this, the various computer networks did not have a standard way to communicate with each other. A new communications protocol was established called Transfer Control Protocol/Internetwork Protocol (TCP/IP). This allowed different kinds of computers on different networks to «talk» to each other. ARPANET and the Defense Data Network officially changed to the TCP/IP standard on January 1, 1983, hence the birth of the Internet. https://clck.ru/3QEh9u

2 Blue Voyant. https://clck.ru/3QEh97

3 Cybercrimes it should be noted are common in the domain of commercial transactions than others.

E-commerce can take the forms of engaging in online shopping, mobile apps conversational commerce via live chat, chatbots, and voice assistants[4]. It can be through Business to Business (B2B), Business to Customer (B2C), Customer to Customer (C2C) and Customer to Business (C2B). It is also defined as the practice of buying and selling goods and services through online consumer services on the internet. It is one of the ways in which companies and individuals carry out business in order to maximize profits within a short time frame while reducing fixed cost over a broad range of assets and transport costs from central business districts[5].

There are several advantages associated to e-commerce, but despite these advantages, cybercrimes have hampered the smooth functioning of e-commerce and have led to loss of profits, trust and confidence amongst business counterparts. There are efforts in Cameroon to combat cybercrimes through legislation and technology but this has not so far seen a significant impact in commercial transactions because of lack of new technologies, weak digital rights management system[6], ease to circumvent technology, convenience, and speed at which cybercrimes are committed. This is more so because of the benefits and satisfaction that perpetrators acquire by committing cybercrimes in e-commerce transactions the very technologies meant for protection.

Cybercrimes and related offences in commercial transactions have in the past years experienced a rise in Cameroon due to an increase in the proliferation of the internet, low cost of procuring electronic devices, global economic crisis, lockdowns due to the COVID-19 pandemic coupled with the introduction of numerous social media platforms. Cybercrimes became popular in Cameroon around 2005 before the enactment of the 2010 laws on cyber criminality and electronic commerce. This can be demonstrated by the earlier cases of; The People v. Obi Roland[7], The People v. Nfang Macknight[8], The People v. Mbah Valery[9], The people of Cameroon v. Tamukum Fonjiyang Ferdinand & one other[10],

4 VentureBeat. (2025, March 15). How to prepare your products and brand for conversational commerce. https://clck.ru/3QEhWE

5 This can be explained by the Von Thunen Theory of agriculture that was developed by Johann Heinrich in 1826. This model predicts human behaviour in terms of landscape and economy based on meticulous mathematical calculations and observations. It explains transport cost from the geographical location of farms to a central business district (CBD). Although it is a theory in agriculture, it relates to e-ecommerce today because the principal objective of e-commerce is to minimize transaction costs from where goods and service originate to where a consumer is located.

6 It refers to technical ways of securing data with the use of passwords, cryptography and steganography.

7 CFIB/55C/2008(Unreported).

8 CFIB/76C/2009(Unreported).

9 CFIB/255/2010(Unreported).

10 (2014) 2 SLR.

The people of Cameroon & another v. Tita Njina Kevin Ndango[11]. All these cases preceded the 2010 Laws on cyber criminality and electronic commerce in Cameroon although some of these cases were reported later. It can be observed in these cases that the civil parties were hardly present in court reason why it shall later be observed in this article that this aspect has caused a surge of cybercrimes in Cameroon.

The first legislations on cybercrimes and e-commerce in Cameroon were enacted in December 2010 following frequent occurrences of cyber criminality cases in Cameroon that related mostly to commercial transactions with the use of online platforms through false pretenses[12] in the guise of legitimate business transactions[13]. That is why the cyber criminality law and the electronic commerce laws in Cameroon were enacted in the same year.

This article therefore treats the impact of cybercrime together with e-commerce transactions in Cameroon for this reason. The cyber criminality law in Cameroon provides the legal framework for investigating and prosecuting cybercrimes alongside the Cameroonian Penal Code and the Criminal Procedure Code (CPC). It elaborates on the types of cybercrimes including unlawful interception, hacking, computer related fraud, offences relating to child pornography etc. The law is not only adjectival but also procedural because it provides rules for investigating and prosecuting cybercrimes with international co-operation.

The 2010 law on electronic commerce in Cameroon on the other hand laid the foundation of e-commerce in Cameroon[14]. This was due to the rapid emergence of e-commerce platforms that followed world trends. The law equally provides substantive elements and procedures to investigate and prosecute allegations of malpractices in e-commerce transactions.

Cybercrimes in Cameroon exist not only in e-commerce transactions but also in other criminal acts like defamation, false pretense, theft, cyberstalking and hacking. Law N° 2010/012 of 21 December 2010 relating to cybersecurity and cyber criminality in Cameroon criminalizes cybercrimes jointly with law N° 2016/007 of 12 July 2016 relating to the Penal Code[15]. While Law N° 2010/021 of 21 December 2010 on electronic commerce in Cameroon establishes offences and punishments for violating the provisions of e-commerce rules. The impact of cyber criminality on e-commerce is huge in Cameroon coupled with the fact that a majority of the country's citizens are not computer savvy and

11 (2010) CCLR 1-126

12 Section 318 of Law N°2016/007 of 12 August 2016 on the penal code.

13 Cybercrimes in Cameroon were officially recognized as a legal issue with the enactment of law N°2010/012 of December 2010 which defines cybercrime offences and provided penalties. Before this law, many instances of cybercriminality were addressed under the penal code.

14 Law N° 2010/021 of 21 December 2010.

15 Section 219 of the Penal Code.

cannot afford internet connections to do research on electronic commercial platforms. This has led to some legal problems like breach of confidence and trust, loss of profit margins in both the micro and macro-economic levels, increased bribery and corruption and a general increase in crime wave since the illicit proceeds acquired from such crimes are often used to perpetuate further offences.

1. Statement of the problem

The introduction of e-commerce in Cameroon and the advent of cyber criminality has had devastating consequences on digital based businesses. This has been compounded by the fact that a vast majority of the Cameroonian population live below the minimum wage level coupled with no official home addresses in the country and no social security numbers attributed to citizens. The lack of digital home addresses and social security numbers makes investigations, the execution of summonses and arrests strenuous because deliveries[16] are usually at the post offices or in other locations on call, or payments are posted to bank accounts that cannot be linked to specific addresses of account owners. Loss of confidence and trust is more nowadays in electronic based commercial transactions in Cameroon because many payments are unsecured[17]. There are huge losses in profit margins and an increase crime wave across the country because illicit profits derived from cybercrimes are sometimes used in furtherance of other crimes like drugs abuse, bribery, corruption, crimes of moral integrity, public drunkenness etc. Cybercrimes have also led to high levels of arbitrary arrests and torture especially on the youths without tangible evidence since some police and judicial officers rely on allegations of scamming due to lifestyles even without any official complaint. Most of these officers hardly prosecute cybercrime in e-commerce cases but prefer to enrich themselves by collecting bribes and freeing suspects. This is compounded by the fact that most victims hardly lodge official complains against the perpetrators. There is also a low level of attention and international co-operation to combat e-commerce related cybercrimes in Cameroon due to inadequate technical infrastructures, personal financial gains and proper follow-up. This has led to an increase in the rate of e-commerce related cybercrimes within the country with most offenders preferring to bribe out their way and turning to other victims to recover their losses[18]. There is also a general atmosphere of fear to file official complaints against cybercriminals because of shame and dire consequences since most victims are accomplices with the perpetrators to indulge

[16] Although controlled deliveries can be done at the post offices to apprehend suspects, this is hardly done in Cameroon.

[17] Return policies are highly ineffective in Cameroon and are hardly embodied in commercial contracts. So the aspect of fear creeps into customers' minds.

[18] It can observed that most cyber offenders under e-commerce hardly stop their acts even after enriching themselves. They mostly perfect their skills.

in illegal transactions form the onset. All of these problems have been escalated by lack of sufficient powers attributed to agencies like the National Financial investigation agency (NFIA) and the National Agency for information and communication technologies (NAICT) to prosecute cybercrimes in Cameroon. Their roles are limited to filing complaints, carrying out investigations and making recommendations with no proper follow-up.

2. Theoretical and Conceptual frameworks

This article is based on several theories.

2.1. The utilitarianism theory

This theory was propounded by John Stuart Mill in 1861. He believed that happiness was the only thing humans do and should desire for their own sake. He believed that because happiness is the only intrinsic good, and since more happiness is preferable to less, the goal of ethical life is to maximize happiness. Jeremy Bentham and John Stuart Mill called it 'the principle of utility' or 'the greatest-happiness principle'. In the context of this article, cybercrimes in e-commerce satisfy the cravings, wellbeing and the happiness of the perpetrators. Therefore, the drive to satisfy personal egos and happiness by offenders may make them desire the benefits of defrauding others by way of committing cybercrimes under e-commerce transactions to gain pleasure and satisfaction. This theory aligns with the Resourceful Evaluative Maximizing Model theory that views individuals as rational actors who are always seeking to maximize their own utility or wellbeing within a given set of constraints. It shows that individuals always strive to find the best possible outcome due to constraints in their resources (Wartiovaara, 2011).

2.2. Transaction cost theory

This theory was introduced by John R. Commons in 1931 (Williamson, 2008). It is cost incurred when carrying out trade. These costs are associated to those in running the economic system of a company and the total costs of carrying out a transaction. It also includes the cost of planning, deciding, changing plans, resolving disputes, and after-sales costs. According to the author, the determinants of transaction costs are frequency, specificity, uncertainty, limited rationality, and opportunistic behaviour. One of the objectives of e-commerce is to significantly reduce transaction costs by way of electronic data interchange (EDI). The Transaction Cost Theory (TCT) focuses on minimal effort on resources and the cost required parties to exchange their goods and services. The objective of this theory is to maximize transaction performance while minimizing costs which is the main objective of e-commerce. This theory can be compared

to the Von Thunen model that dwells on transaction cost like transportation of crops from farms to a central business district. The sales price is mostly determined according to the distance between the central business district and where the farm produce is sold. This theory is related to e-commerce because of the distances between where goods and services are located, where negotiations take place and where they are delivered. E-commerce in consideration to this model leads to reduced cost, irrespective of where the goods are manufacture and where they are delivered, as opposed to transactions that are concluded in a manner where the parties have to travel over long distances to carry out negotiations and carry out shopping in real world.

2.3. Rational choice theory

This theory was propounded by Adam Smith in 1776 and later articulated by the sociologist George Homans in 1961. The theory is based on behavioral psychology. The theory involves achieving a goal using the most cost-effective method without reflecting on the worthiness of that goal. The goals may be self-regarding, selfish, or materialistic (Snidal, 2013). The theory gives guidelines that help to understand economic and social behaviour and used in criminology. It helps to predict the outcome and pattern of choice and assumes that individuals are self-interested when decisions are based on optimizing preferences by balancing costs and benefits. This phenomenon is common in e-commerce where many options are available in online shopping where different choices may be made in preference to others. The goals of cybercriminals in e-commerce are usually self-centered without regards to the consequences caused to their victims after depriving them of their wealth. The theory is also related to crime where an individual can decide to commit an offence and be caught or takes risk to commit and offence and go free while benefiting. This is a typical phenomenon in e-commerce related cybercrimes where it is difficulty to catch perpetrators the act.

3. Related Literature

Some authors have written independently in the areas of cybercrime and e-commerce and have made little connection between cybercrimes and e-commerce. While few authors have highlighted the impact of cybercrime on e-commerce transactions, they have not critically examined the relation between the two, which is metaphoric. This is because the same sanctions that exists in conventional crimes under commercial transactions exist under cybercrimes in electronic commerce transactions. What happens offline is the same as what happens online with the difference being the mode and interface used in carrying out the same act with the same act.

Reyns et al. (2011) dwells on the fears caused by cybercrimes because of victimization. The author examines the relationship between risk in cyberspace for fear of being a victim

of cybercrime. His analysis is based on information collected from undergraduate students at the University of Cincinnati. The major finding in his research shows that many people are worried to become victims of cybercrimes. He further emphasizes on the category of offenders and their behavioral pattens in relation to status and gender. His work examines behavioural frequencies that have great effects on the levels of fear because of cybercrime victimization. He agrees that fear and victimization in cyberspace are based on perceived risks. His work doesn't analyze how these fears can be allayed to encourage e-commerce.

Böhme и Moore (2012) dwell on cybercrimes in online shopping and how to prevent them. Their main finding was that cybercrimes have led to the reduction in the rate of transactions like; online banking, online shopping which has caused huge negative effects. The paper concluded that people who do not know about cybercrimes are more likely to engage in e-commerce transactions like online shopping. Their work is mostly limited to online shopping meanwhile e-commerce is broader.

Y. Abubakari (2020) demonstrates how people lose many opportunities for fear of being scammed. The author shows the reasons, impacts and limitations of cybercrime policies in Anglophone West Africa. He also demonstrates how cybercrime perpetrators lose focus in education and that the reason for the growth of cybercrimes is associated with economic strains and corruption at the governmental level. The author considers hindrances in cybercrime policy because of corruption, government interference, ineffective implementation of cybercrime laws and inconsistencies in the content of cybercrime policies. His research focused on Ghana, Nigeria, and Sub-Saharan Africa as a representative sample for Anglophone West Africa because the prevalence of internet fraud in West Africa is centered around the Anglophone West African counties like; Ghana, Nigeria, Liberia, Sierra Leone, Gambia and part of Cameroon, with Nigeria and Ghana being the most notorious.

André Boraine and Ngaundje Leno Doris (2019) wrote on the fight against cybercrime in Cameroon. They focused mainly on the conflict in the Anglophone regions of Cameroon and showed a link between the conflict and the increased rate of cybercrimes due to the conflict. They examine the role of the government of Cameroon in the fight against cybercrimes and analyzed some of the legal provisions used to combat cybercrimes in Cameroon. Their paper examined why cybercrimes are prevalent in Cameroon and recommend measures that can be put in place to combat cybercrimes in Cameroon. They did not show the direct impact of cybercrimes on e-commerce transactions in Cameroon. Their paper raises awareness and contributes to knowledge in data protection rules, especially among investigating officers, students, specialists, and non-specialist legal practitioners.

Most available literature as examined above is focused on either cybercrime or e-commerce. This article shows the direct impacts that cybercrimes have on e-commerce transactions in Cameroon and makes recommendations.

4. Common types of e-commerce related cybercrimes in Cameroon

Cybercrimes in e-commerce transactions in Cameroon caused an approximate loss of 12.2. billion francs CFA in 2021 with scamming and phishing accounting for approximately half the amount[19]. There are different types of e-commerce related cybercrimes that affect Cameroon and the world. It may be observed that the nature of cyberspace communications does not limit cybercrimes to particular geographical territories. A person may be in a country and commit cybercrime across different countries with different systems of law. Therefore, there is need for international co-operation to combat cybercrimes, especially e-commerce related cybercrimes. This same reason accounts for the absence of complainants in legal proceedings and the growth of cybercrimes[20]. The types of cybercrimes associated to e-commerce include but not limited to: spamming, salami attacks, virus bombarding, cyber defamation etc. The common types in Cameroon are scamming[21], phishing[22] and bank card skimming[23]. Most of these crimes in Cameroon are oriented towards financial gains and occur in e-commerce at inter personal level.

4.1. Scamming

Scamming originates from the word scam. It is a dishonest plan to make money or getting an advantage especially by tricking people. It becomes a scheme if the plan is in a large scale. Scheming is relatively rare in Cameroon[24]. Scamming is also a confidence trick to defraud a person or group after gaining their trust by taking advantage of a combination of factors like the victim's naivety, compassion, vanity, confidence and greed (Orbach & Huang, 2018). In Cameroon, according to a 2021 report of the Ministry of Post and Telecommunications (MINPOST), the rate of scamming was 60 % in Yaounde, Douala, Buea and Noun amongst unemployed young people aged between 16 and 35 years[25].

Scamming is not a new phenomenon but has grown with the proliferation of ICT tools and the internet. In ancient Greece, cups and balls trick were used as forms of deception and in same Greece, a 'confidence man' called Thompson who was a swindler asked

---

19 2021 report of the Ministry of Post and Telecommunications. Note 30. https://clck.ru/3QMjCH

20 One of the reasons behind lack of will in Cameroon to prosecute cybercrimes is because the victims are sometimes citizens of foreign countries and lack interest to prosecute in Cameroon due to some legal challenges, cost and fear. This aspect permit bribery and corruption since the victims are not usually available in Cameroon.

21 It is false pretence with the use of electronic communication protocols.

22 To behave as a trusted person to gain access to sensitive information.

23 A method to obtain bank card information while they are used on an automatic teller machine.

24 2021 report of the Ministry of Post and Telecommunications. Note 6 p. 1346. https://clck.ru/3QMjCH

25 Ibid.

his victims to express confidence in him by giving him money rather than gaining their confidence in a more nuanced way. He was not successful and was arrested in July 1849[26]. E-commerce related cybercrimes became noticeable in Cameroon in 2005 before the enactment of the 2010 law on cybersecurity. Before then, courts relied mostly on the Penal Code[27] to adjudicate such cases as was seen in the case of The People v. Obi Roland[28]. This was a case of scamming that was heard by the Court of First Instance Buea in 2008 before the enactment of the 2010 laws on cybersecurity and electronic commerce law. Section 318 of the Cameroonian Penal Code was used as basis of the judgement. The accused was found guilty and sentenced to six years imprisonment. This was the same position held by thesame court in The People v. Nfang Macknight[29], where the accused was found guilty by the same court for similar reasons.

Even though the courts have often found the accused persons guilty, surprisingly many of such cases hardly get to Cameroonian courts. The procedure in determining cyber cases is often riddled with incompatibilities like the violation of rights under sections 3 and 8 of the CPC that have led to the discharge of some accused persons as was seen in the cases of The People v. Mbah Valery[30], The people of Cameroon v. Tamukum Fonjiyang Ferdinand & one other[31]. It can be observed that because of the high level of bribery and corruption in e-commerce related cybercrime cases like scamming, courts are hardly seized nowadays in Cameroon to adjudicate on such cases. The scammers and investigating officers prefer to collude to bribes for their freedom. This may be however difficult in situations where a complainant decides to follow up prosecution by filling a civil claim before a competent court. This was so in the case of The people of Cameroon & another v. Tita Njina Kevin Ndango (Jansson & von Solms, 2011) where the Court of First Instance Buea found the accused guilty of all the counts in the charge because the complainant travelled from Switzerland to attend the hearing.

4.2. Phishing

It refers to a type of scamming where victims are tricked to reveal sensitive information or by installing a malware that may contain salami attacks, viruses, worms, that can mirror a targeted website (Jansson & von Solms, 2011). It may be done in a manner that there is a slight alteration in the spelling or numerals of a website to resemble the original website such that a victim may not know that he is dealing with a fake website.

---

26 2021 report of the Ministry of Post and Telecommunications. Note 29. https://clck.ru/3QMjCH

27 Law N°2016/007 of 12 July 2016 on the penal code of Cameroon.

28 CFIB/55C/2008(Unreported).

29 CFIB/76C/2009(Unreported).

30 CFIB/255/2010(Unreported).

31 (2014) 2 SLR.

Phishing is commonly used in Cameroon on mobile networks where victims are called or sent text messages from a cell phone or smartphone to deliver a bait message. The fake messages are usually on grounds of erroneous financial deposits into the victims' mobile money accounts. The perpetrator meanwhile has logged the telephone number of the victim into the mobile operator's version of desktop application (APP). The perpetrator then tricks the victim to verify his mobile account details by entering his passcode for a refund with a promise of reward. If the victim request a mobile money service on his account and enters his passcode on his device, it is simultaneously communicated to the perpetrator's desktop app where he gains access to withdraw or transfer funds. This aspect is notorious in Cameroon and almost all mobile telephone users in Cameroon have experienced this phenomenon with a majority falling for the trick. This aspect not only generate fear but also distorts the idea of e-commerce carried out by mobile telephone companies as one of the ways that the telephone operators use to conduct their businesses online. Some cases of this nature have been successful before the laws courts because it is possible to track the number of the perpetrator although, the criminals are at times smart enough to register the withdrawing account number in the names of different individuals who are based in different geo-locations.

4.3. Bank Card Skimming

This is when technology devices are installed on or inside automatic teller machines or at other sales points where perpetrators can capture card details and replicate them[32]. It is also a tactic that criminals use to obtain sensitive information from a debit or credit card. These devices capture and store data that fraudsters later use to make purchases or withdrawals at different times[33]. Bank card skimming is not very common in Cameroon because of the scarcity of the technological devices used and because most ATMs and sales points are secured with cameras and guards. Although it is still possible for skimmers to physically steal card details when they are left carelessly.

4.4. Socio-legal impacts of cybercrimes on e-commerce transactions in Cameroon

Cybercrimes have many negative consequences on businesses and e-commerce users ranging from loss of confidence to bankruptcy of businesses and other social ills across the countries (Luu et al., 2025; Higgs & Flowerday, 2025b; Lee et al., 2023; Holt, 2022; Hornuf et al., 2025). Cybercrimes related to e-commerce has social and legal consequences and creates the fear of victimization on potential users willing to pay

[32] Federal Bureau of Investigation. Skimming. https://clck.ru/3QEhzV

[33] BrightBridge Credit Union. https://clck.ru/3QEi7p

for goods and services through online platforms. Fright is an element that affects the physical and psychological dispositions of every human being. It erodes confidence and pushes people to do cost benefit analysis with some results that lead people to prefer cash payments on the spot[34].

The proliferation of cybercrimes in e-commerce transactions can be explained under the utilitarian and the rational choice theories[35] where human nature tends to situate people in a position that maximizes their satisfaction even if it means to engage in irrational choices that at times may achieve a positive goal without fear of the consequences. The utilitarian view considers satisfaction as a factor that can overshadow rational behavior. This explains why a majority of youths and young people in Cameroon in some places like Buea and Bamenda have picked up scamming as a trade within the Molyko and Bambili neighborhoods respectively. E-commerce related cybercrimes have also grown rapidly in these areas because of the socio-political conflict in the English-speaking regions of Cameroon. This situation is further aggravated by a high unemployment rate amongst the youths, a low minimum wage level across the country[36] and a general low salary scale in Cameroon. Most of the youths in these areas in Cameroon especially the females open fake online shops purporting to do e-commercial transactions through social media platforms to mask the illicit sources of their income. E-commerce related cybercrimes were mostly committed in the past by males who engage in deceptive tricks by electronically producing rare images of objects and animals that do not exist to entice their victims.

Due to the introduction of modern ICT infrastructures and easy access to the internet at affordable costs in Cameroon, Scamming and phishing are nowadays common in most populated cities. A survey carried out by the author between January to April 2025 at Molyko in Buea where 250 sample responses were obtained from random participants revealed that at least 60 % of the youths in Molyko are involved in e-commerce related cybercrimes. The data collected also revealed that 10 % of the perpetrators were females while 50 % were males ranging from the ages of 16 to 35 years. It was also observed that most of the victims are people residing in Molyko and know their victims. A few victims were unknown by the perpetrators while some of the victims resided in different cities of Cameroon and abroad.

---

34 Many purchasers in Cameroon prefer to buy and do spot payments in cash for products that they can see and fill. This is the reason why most people in Cameroon move with cash despite the attendant risks.

35 See the Rational theory supra.

36 The minimum wage in Cameroon is 46,939frsCFA less than what obtains in other neigbhouring African countries.

The survey also found out that the illicit financial benefits derived from these cybercrimes vary significantly in amounts with few perpetrators making huge amounts almost on regular basis while some just barely make pocket money for the day. The activities of the perpetrators involved mostly scamming and phishing. The survey concluded that most of the youths in Molyko get involved because of the need to satisfy their ego for money, competition, peer pressure, strain and the insignificant number of prosecutions because of corrupt practices. It was also concluded that most perpetrators are still actively involved in cybercrimes related to e-commerce transactions while still perfecting their skills.

It was also discovered that many youths have dropped out of school as observed by Yushawu Abubakari in his article in 2020 (Abubakari, 2020). While some youths who have been submerged by cybercrimes paid others out of their illicit gains to sit for their school examinations[37].

Occult practices[38] are also common with dire consequences amongst cybercriminals in Cameroon because most youths with low intellectual capacity are lured to believe that their successes are dependent on such practices to convince their victims. These practices have led to some fatalities and the emergence of organized gangs of fraudsters who at times jointly contribute bribes to 'make a way out' for their counterparts in trouble. The levels of these occult practices extend to other social aspects like homosexuality, Lesbianism incest and sexual activities under the same roof[39].

All these factors as earlier observed have contributed to a high level of school dropouts thereby increasng the level of illiteracy amongst the youths in Cameroon. Meanwhile the Cameroonian government has despite this sacrificed to educate its youths by subsidizing primary, secondary and university education throughout the country.

4.5. The Proliferation of other offences due to e-commerce related cybercrimes

Illicit wealth derived from cybercrimes related to e-commerce in Cameroon has led to the proliferation of other offences like drugs abuse, prostitution, public drunkenness, impersonation in examinations, sexual offences, defamation, violence, corruption, identity theft, assault, battery reckless driving etc. Unexpected wealth can lead to psychopathic tendencies especially amongst youths whose mental capacities are still developing. This was also observed in the survey conducted.

---

37 Some of the cases of impersonation have been detected during disciplinary board sessions by authorities of universities based in Molyko.

38 Occultism describes various practices and beliefs related to the study of manipulation of supernatural forces. It involves a wide range of practices including divination, magic, alchemy, astrology and spiritualism.

39 Although there are human rights activists today in Cameroon who carry out the activiies aimed at protecting the rights of lesbians, gays, bisexuals, transgender and queer people (the international movement of LGBTQ is recognized as extremist and banned in the territory of the Russian Federation).

4.5.1. Loss of trust and confidence by e-commerce users

It is common today to see most youths in Cameroon who are living a flamboyant lifestyle while riding cars of a particular mark without any proof of their financial means. This attitude is directly linked to the high rate of e-commerce related cybercrimes and corrupt practices. Cameroon's corruption index according to transparency international in 2024 was 26 points /100, far higher than most countries in the world[40]. These results can be partially attributed to cybercrimes in e-commerce and the accompanying frivolous investigative practices that are laden with bribery and corruption with little follow up at the level of the International Police Organization (INTERPOL)

The Cameroonian 2010 law on cybersecurity attributes the jurisdiction and authority of investigative officers[41] who should abide by the rules of criminal procedure as captured by the 2005 Criminal Procedure Code[42]. But unfortunately, most of these officers collude with the perpetrators for personal gains and so most cybercrime cases hardly go to court for proper determination[43]. It is common to find cases where a suspected cyber offender is arrested because of his suspicious financial activities by police and gendarme officers even without a complaint, who accompany the suspect to ATM machines to collect their own purported share of ill-gotten money. These offices at times force the suspects to transfer money from their mobile money accounts to their own phones before letting them go. It is common to see the accomplices of suspects contributing bribes for the release of one of theirs in trouble. Some cyber offenders have made it a habit to put some officers on their pay roll who have always shielded them from any possible eventuality. Suspects who fail to co-operate with the officers have more often been subjected to arbitrary arrest and detention with torture. This raises an issue of due process and failure by the officers to maintain the rule against torture as enshrined in the ICCPR, ACHPR and other national legislations.

All these negative aspects put together has eroded trust and confidence on the country's citizens to a greater extent both home and abroad. Many opportunities to carry out legitimate e-commerce transactions in Cameroon with foreigners has been compromised for lack of trust. It may be noted that many Cameroonians residing abroad are well known for cybercrimes that has led to the incarcerations with some serving long prison sentences in foreign countries. The phenomenon of hiding cyber criminals for a bribe is experienced today in many African countries (Sarefo et al., 2023; Matias, 2025). Most investigative officers prefer to take a bribe than to prosecute perpetrators before the law courts because of their ego.

---

40 Transparency International. https://clck.ru/3QEjAB

41 Section 52(1) of Law N° 2010/012 of 21 December 2010 relating to cybersecurity and cybercriminality in Cameroon

42 Sections 59 and 60 of Law N°2005 of 27 July 2005 on the Criminal Procedure Code.

43 Some of the bribes are taken most often from the suspects by force and coercion.

Loss of trust and confidence in commercial transactions is experienced at the interpersonal levels of relationships, corporate bodies and sovereign states (Wright & Kumar, 2023; Yi et al., 2024; Porcedda, 2023; Sarkar & Shukla, 2024; Tok et al., 2025; Onwuadiamu, 2025). As noted by B. Rainer and Tyler Moore (2012), most individuals who have been victims of cybercrimes in online commercial transactions end up never engaging in it again for fear of victimization while those who do not know about cybercrimes are more likely to engage in online commercial transactions. The impact of cybercrimes on e-commerce has therefore discouraged many foreigners from engaging in e-commerce transactions with Cameroonians in general. This loss of confidence is linked not only to Cameroonians but also to the neighboring West African citizens, as examined by Yushawu in his paper (Abubakari, 2020).

Significant financial losses have been experienced in Cameroon due to e-commerce related cybercrimes because there is a great latitude of choice and freedom in cyberspace[44], prospective e-commerce users prefer to deal with other nationals for fear of being victimized on Cameroonian e-commerce platforms. The estimated losses on the economy of Cameroon caused by e-commerce related cybercrimes in 2021 was revealed by the Ministry of Post and Telecommunication which is the competent ministry that deals with communication issues in Cameroon. The ministry's report estimated according to ANTIC and ANIF[45] that Cameroon lost approximately 12.2 billion francs CFA in 2021 due to scamming and phishing.

4.5.2. Arbitrary arrests and detentions due to e-commerce related cybercrimes

Arbitrary arrests and detention for suspicion on e-commerce related cybercrimes without any complain in violation of due process are frequent in the major cities of the Northwest and Southwest regions of Cameroon with Buea, Bamenda and Limbe being notorious. The preamble of the Cameroon constitution[46] is clear that no person shall be arrested or detained except in the manner determined by law. The law that determines such arrest in Cameroon is the Criminal Procedure Code[47]. Contrary to the constitutional and procedural provisions on arrest and detention, most cybercrime suspects are arbitrarily arrested without arrest warrants nor under the flagrante delicto procedure[48] as prescribed by the CPC. This can be attributed to the fact that the persons conducting such arrest are aware that they could make quick income through bribes that some cybercriminals are willing to pay as trade-off for their freedom.

---

44 Lawrence Lessig puts it as democracy in cyberspace.

45 2021 report of the Ministry of Post and Telecommunications. Note 24 and 25. https://clck.ru/3QMjCH

46 Law N° 96-6 of 18 January 1996 (as revised)

47 Section 30 of Law N° 2005 of 27 July 2005 on the Criminal Procedure code.

48 Section 31.

Due process is a fundamental aspect of legal proceedings that seeks to protect civil rights. Breaching it may lead to the nullification of an entire case[49]. In the cases of The People v. Mbah Valery[50], and the people of Cameroon v. Tamukum Fonjiyang Ferdinand & one other[51], the Court of first Instance Buea acquitted and discharged the accused persons for violating some procedural aspects of criminal procedure that included illegal arrests and detention. The trial court held in the latter case that although the character of the accused persons was doubtful, they should walk away from the court as free people. This was because of serious procedural errors that violated the provisions of sections 3 and 8 of the Criminal Procedure Code. The decision was not meant to encourage the actions of the accused but went a long way to show the importance of protecting fundamental human rights[52]. It therefore becomes problematic when suspects are intercepted by the forces of law and order in violation of their constitutional and legal rights under the pretext of enforcing the law. Most of the victims of such arrest are not only tortured at times but their financial resources are also extorted by coercion. This justifies the purported "good will" of some police officers to fight cybercrimes in Cameroon.

Arbitrary arrests and detention without due process ought to be discouraged because it violates the fundamental principles of civil liberty enshrined in most international conventions and domestic laws. Article 9 of the International Covenant on Civil and Political Rights (ICCPR) 1966 prohibits arbitrary arrest and detention while article 6 of the African Charter on Human and People's Rights (ACHPR)1987 guarantees the right to liberty and clearly articulates that the deprivation of this freedom must be for reasons and conditions laid down by the law. These legal provisions are applicable in Cameroon pursuant to article 45 of the 1996 constitution as revised[53].

Cases of illegal arrests and detentions due to e-commerce related cybercrimes are common in the Northwest and Southwest regions of Cameroon as compared to other regions of the country (Abubakari, 2020). This has partially been attributed to the socio-political crisis plaguing the two regions (Boraine & Leno Doris, 2019). The crisis has contributed to an increase in e-commerce related cybercrimes meanwhile the officers who have been deployed to these regions have seized the opportunity to unlawfully arrest and detain suspect for their personal benefits under the pretext of economic crimes. This has invariably led to a fragrant disregard of the legal provisions that guarantee civil liberties.

---

49 See sections 3 and 8 of the CPC dealing with absolute and relative nullity. See also the decision of Lord Denning in the case of United Africa Company Limited (U.A.C) v. Macfoy dealing on nullity.

50 CFIB/255/2010(Unreported).

51 (2014) 2 SLR.

52 See ICCPR and the ACHPR.

53 Article 45 of Law N° 96-6 of 18 January 1996 on the Cameroonian Constitution.

4.6. Economic impacts on e-commerce caused by cybercrimes

The growth of e-commerce related cybercrimes in Cameroon has led to financial losses and profits on individuals and corporate bodies, leading low incentives to engage in e-commerce transactions. This is because fears caused by cybercrimes due to victimization have a greater impact in cyberspace because many people are worried of becoming victims of cybercrimes (Yokotani & Takano, 2021). This invariably has a nexus with the demand for e-commerce transactions, and may significantly reduce turnover and output because of loss capital and profits due the actions of cybercriminals. Some financial losses because of e-commerce related cybercrimes may lead to bankruptcy. The amount of financial loses for instance as observed above cause due to cybercrimes in 2024 was about 12.2 billion francs CFA. This has a huge impact on the economy and besides the illicit financial gains made from e-commerce related cybercrimes are not taxable.

All of the above distorts the smooth functioning of e-commerce in Cameroon. One of the purposes of e-commerce is to achieve fast turnover. With the proliferation of cybercrimes in Cameroon, there is a huge challenge as the demand for e-commerce services dwindle. Loss of trust and confidence due to cybercrimes in e-commerce has also significantly destroyed the economy of the country in different ways increasing inflation[54], money laundering and currency counterfeiting.

Conclusion

This article has examined the impacts of Cybercrimes related offences on e-commerce in Cameroon and has observed that most cybercrimes in Cameroon are targeted towards fake e-commerce transactions. It has been found out that the economic situation of Cameroon, the easy access to ICT tools, the conflict in the Southwest and North west regions of Cameroon and the need to live flamboyant live styles are some of the contributing factors in Cameroon that have led a high rate of e-commerce related cybercrimes. It was found out that there are socio-legal and economic impacts and the proliferation of other crimes as a result. The high rate of cybercrimes related to e-commerce transactions in some major cities of Cameroon are mostly perpetrated by youths with the number of potential future offenders on the rise. This was demonstrated by the results of a survey carried out and the reports of the Ministry of Post and Telecommunication. It was also discovered that the role of the institutions charged with combatting cybercrimes in Cameroon is limited. The findings also reveals that e-commerce related cybercrimes in Cameroon has led to loss of trust and confidence with arbitrary arrests and detentions in violation of some international treaties, constitutional and criminal procedure provisions. It has

[54] It can be observed that the prices of basic commodities in Buea is high as compared to other cities in cameroon because of illegal wealth that is used in Buea by mostly youths who are scammers.

been observed in this article that most suspect are hardly brought before the competent law courts because police officers prefer to take bribes and set the suspects free. The findings further show that cybercrimes in e-commerce related transactions has tarnished the reputation of Cameroonians at home and abroad and most foreigners have lost trust and confidence to conduct online businesses with Cameroonians. This loss of trust and confidence is attributed to lack of proper home address systems and social security numbers attributed to Cameroonians.

It is therefore recommended that legal, technological, social and economic reforms should be instituted in Cameroon to resolve the issues identified. The legislations in force should be improved upon by specifically addressing the procedural rules to follow at the investigative and litigation stages on cases of e-commerce related cybercrimes. The law should empower the National Financial investigation agency (NFIA) and the National Agency for information and communication technologies (NAICT) to prosecute cases of e-commerce related cybercrimes. The forces of law and order should be trained regularly on arrest and detention on cybercriminal while providing them with incentives for successful e-commerce related cybercrime prosecutions. Serious sanctions should also be applied in cases of corruption and bribery under cybercrimes. The minimum wage level and the salary scale of Cameroon should be improved while creating new jobs to absorb idle youths. Schools should introduce cybersecurity and e-commerce lessons in their curriculum from the elementary level up to the universities. A proper home address system and social security numbers should be instituted in Cameroon for easy identification. Digital rights management systems and tracking technology should be improved.

References

Abubakari, Y. (2020). The reasons, impacts and limitations of cybercrime policies in Anglophone West Africa: A review. *Social Space Journal*, *12*(3), 45–67.

Alhadidi, I., Nweiran, A., & Hilal, G. (2024). The influence of Cybercrime and legal awareness on the behavior of university of Jordan students. *Heliyon*, *10*(12), e32371. https://doi.org/10.1016/j.heliyon.2024.e32371

Arroyabe, M. F., Arranz, C. F., De Arroyabe, I. F., & De Arroyabe, J. C. F. (2024). Revealing the realities of cybercrime in small and medium enterprises: Understanding fear and taxonomic perspectives. *Computers & Security*, *141*, 103826. https://doi.org/10.1016/j.cose.2024.103826

Böhme, R., & Moore, T. (2012). The market for stolen data: Estimating the costs of data breaches. *IEEE Security & Privacy*, *10*(2), 43–49. https://doi.org/10.1109/MSP.2012.40

Boraine, P. A., & Leno Doris, D. N. (2019). The Fight against Cybercrime in Cameroon. *International Journal of Computer (IJC)*, *35*(1), 87–100. https://doi.org/10.53896/ijc.v35i1.1469

Chris, H. et al. (2005). *Criminology*. Oxford: Oxford University Press.

Edwards, M., & Hollely, N. M. (2023). Online sextortion: Characteristics of offences from a decade of community reporting. *Journal of Economic Criminology*, *2*, 100038. https://doi.org/10.1016/j.jeconc.2023.100038

Garner, B. A. (1999). *Black's Law Dictionary* (7th ed.). Minnesota, West Group.

Gupta, A., Matta, P., & Pant, B. (2025). Analyzing the social conduct of users to detect and prevent cybercrime across social networks. *Procedia Computer Science*, *258*, 4269–4278. https://doi.org/10.1016/j.procs.2025.04.676

Higgs, J., & Flowerday, S. (2025a). Detecting cybercrime in online video gaming. *Computers & Security*, *156*, 104528. https://doi.org/10.1016/j.cose.2025.104528

Higgs, J., & Flowerday, S. (2025b). Cybercrime and virtual laundering in video game marketplaces: A Bayesian P-inductive argument. *Technology in Society*, *83*, 102978. https://doi.org/10.1016/j.techsoc.2025.102978

Holt, T. J. (2022). Understanding the state of criminological scholarship on cybercrimes. *Computers in Human Behavior*, *139*, 107493. https://doi.org/10.1016/j.chb.2022.107493

Hornuf, L., Momtaz, P. P., Nam, R. J., & Yuan, Y. (2025). Cybercrime on the ethereum blockchain. *Journal of Banking & Finance*, *175*, *107419*. https://doi.org/10.1016/j.jbankfin.2025.107419

Jansson, K., & von Solms, R. (2011). Phishing for phishing awareness. *Behaviour & Information Technology*, *32*(6), 584–593. https://doi.org/10.1080/0144929x.2011.632650

Lee, S., Kang, I., & Kim, H. (2023). Understanding cybercrime from a criminal's perspective: Why and how suspects commit cybercrimes? *Technology in Society*, *75*, 102361. https://doi.org/10.1016/j.techsoc.2023.102361

Luu, T. J., Samuel, B. M., Jones, M., & Barnes, J. (2025). Exploring how the Dark Triad shapes cybercrime responses. *Personality and Individual Differences*, *244*, 113250. https://doi.org/10.1016/j.paid.2025.113250

Matias, C. F. F. (2025). Access revisited: AI training at the intersection of copyright and cybercrime laws. *Computer Law & Security Review*, *57*, 106149. https://doi.org/10.1016/j.clsr.2025.106149

Onwuadiamu, G. (2025). Cybercrime in Criminology; A Systematic Review of criminological theories, methods, and concepts. *Journal of Economic Criminology*, *8*, 100136. https://doi.org/10.1016/j.jeconc.2025.100136

Orbach, B., & Huang, L. (2018). Con men and their enablers: The anatomy of confidence games. *Social Research: An International Quarterly*, *85*(4), 795–822. https://doi.org/10.1353/sor.2018.0050

Payne, B. K. (2020). *The Handbook of International Cybercrime and Cyberdevidence*. London: Palgrave Macmillian. https://doi.org/10.1007/978-3-319-78440-3

Porcedda, M. G. (2023). Sentencing data-driven cybercrime. How data crime with cascading effects is tackled by UK courts. *Computer Law & Security Review*, *48*, 105793. https://doi.org/10.1016/j.clsr.2023.105793

Reyns, B. W., Henson, B., & Fisher, B. S. (2011). Being Pursued Online: Applying Cyberlifestyle–Routine Activities Theory to Cyberstalking Victimization. *Criminal Justice and Behavior*, *38*(11), 1149–1169. https://doi.org/10.1177/0093854811421448

Sarefo, S., Mphago, B., & Dawson, M. (2023). An analysis of Botswana's cybercrime legislation. *Procedia Computer Science*, *219*, 1023–1033. https://doi.org/10.1016/j.procs.2023.01.380

Sarkar, G., & Shukla, S. K. (2024). Reconceptualizing online offenses: A framework for distinguishing cybercrime, cyberattacks, and cyberterrorism in the Indian legal context. *Journal of Economic Criminology*, *4*, 100063. https://doi.org/10.1016/j.jeconc.2024.100063

Snidal, D. (2013). Rational Choice and Interntional Relations. In *Handbook of International Relations*. London, Sage. https://doi.org/10.4135/9781446247587.n4

Tok, Y. C., Zheng, D. Y., & Chattopadhyay, S. (2025). A Smart City Infrastructure ontology for threats, cybercrime, and digital forensic investigation. *Forensic Science International Digital Investigation*, *52*, 301883. https://doi.org/10.1016/j.fsidi.2025.301883

Wartiovaara, M. (2011). Rationality, REMM, and Individual Value Creation. *Journal of business Ethics*, *98*(4), 641–648. https://doi.org/10.1007/s10551-010-0643-6

Williamson, O. E. (2008). Outsourcing: Transaction Cost Economics and Supply Chain Management. *Journal of Supply Chain Management*, *44*, 2–82. https://doi.org/10.1111/j.1745-493x.2008.00051.x

Wright, D., & Kumar, R. (2023). Assessing the socio-economic impacts of cybercrime. *Societal Impacts*, *1*(1–2), 100013. https://doi.org/10.1016/j.socimp.2023.100013

Yi, M., Li, J., & Shen, G. (2024). Can targeted poverty alleviation reduce criminal offenses? Empirical evidence from China judgments online data. *China Economic Quarterly International*, *4*(1), 55–68. https://doi.org/10.1016/j.ceqi.2024.03.003

Yokotani, K., & Takano, M. (2021). Predicting cyber offenders and victims and their offense and damage time from routine chat times and online social network activities. *Computers in Human Behavior*, *128*, 107099. https://doi.org/10.1016/j.chb.2021.107099

Yokotani, K., & Takano, M. (2022). Effects of suspensions on offences and damage of suspended offenders and their peers on an online chat platform. *Telematics and Informatics*, *68*, 101776. https://doi.org/10.1016/j.tele.2022.101776

Author information

Lekunze Anoumbuandem Benvolio – PhD, Lecturer, Department of English Law, University of Buea
Address: PO Box 63, Buea, Cameroon
E-mail: benleku@yahoo.com
ORCID ID: https://orcid.org/0009-0005-9947-0639

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – April 27, 2025
Date of approval – May 9, 2025
Date of acceptance – September 25, 2025
Date of online placement – September 30, 2025

Научная статья
УДК 34:004:343.721:004.8
EDN: https://elibrary.ru/tnqlxy
DOI: https://doi.org/10.21202/jdtl.2025.21

Влияние киберпреступности на электронную коммерцию в Камеруне: социально-правовые и экономические последствия

Анумбуандем Бенволио Лекунзе

Университет Буэа, Буэа, Камерун

Ключевые слова

безопасность,
Камерун,
киберпреступность,
мошенничество,
право,
правосудие,
транзакции,
фишинг,
цифровые технологии,
электронная коммерция

Аннотация

Цель: проанализировать влияние киберпреступности на операции электронной коммерции в Камеруне и оценить эффективность существующих правовых механизмов противодействия киберугрозам.

Методы: исследование базируется на теориях утилитаризма, транзакционных издержек и рационального выбора. Применена методология качественного исследования с использованием доктринального метода. Проведен комплексный анализ правовых актов Камеруна в сфере кибербезопасности и электронной коммерции. Выполнено социологическое обследование с получением 250 выборочных ответов от жителей района Молико в городе Буэа в период с января по апрель 2025 г. Исследованы судебные прецеденты и статистические данные Министерства почты и телекоммуникаций Камеруна.

Результаты: установлено, что киберпреступления привели к потере доверия к операциям электронной коммерции в Камеруне, что отражается на снижении желания граждан осуществлять онлайн-транзакции. Выявлено, что более 60 % молодежи в возрасте от 16 до 35 лет в крупных городах Камеруна либо вовлечены в киберпреступления, связанные с электронной коммерцией, либо пострадали от них. Зафиксирован рост числа женщин среди киберпреступников. Определены основные виды киберпреступлений: мошенничество, фишинг и хищение средств с банковских карт.

Научная новизна: комплексный междисциплинарный анализ влияния киберпреступности на электронную коммерцию в контексте развивающейся африканской экономики. Впервые проведено эмпирическое исследование масштабов киберпреступности в конкретном регионе Камеруна с количественной оценкой вовлеченности молодежи в противоправную деятельность. Разработана теоретическая модель, объединяющая концепции утилитаризма, транзакционных издержек и рационального выбора для объяснения мотивации

киберпреступников. Выявлены специфические социально-правовые факторы, способствующие росту киберпреступности в условиях социально-политического кризиса.

Практическая значимость: результаты исследования имеют важное прикладное значение для совершенствования правовых, технологических, социальных и экономических механизмов противодействия киберпреступности в Камеруне. Предложенные рекомендации включают реформирование процессуального законодательства, расширение полномочий специализированных органов, введение системы домашних адресов и номеров социального страхования, повышение минимальной заработной платы и интеграцию курсов кибербезопасности в образовательные программы. Полученные данные могут быть использованы правительственными структурами, судебной системой, образовательными учреждениями и международными организациями для разработки эффективных стратегий борьбы с киберпреступностью и развития безопасной цифровой экономики.

Для цитирования

Лекунзе, А. Б. (2025). Влияние киберпреступности на электронную коммерцию в Камеруне: социально-правовые и экономические последствия. *Journal of Digital Technologies and Law*, *3*(3), 512–536. https://doi.org/10.21202/jdtl.2025.21

Список литературы

Abubakari, Y. (2020). The reasons, impacts and limitations of cybercrime policies in Anglophone West Africa: A review. *Social Space Journal*, *12*(3), 45–67.

Alhadidi, I., Nweiran, A., & Hilal, G. (2024). The influence of Cybercrime and legal awareness on the behavior of university of Jordan students. *Heliyon*, *10*(12), e32371. https://doi.org/10.1016/j.heliyon.2024.e32371

Arroyabe, M. F., Arranz, C. F., De Arroyabe, I. F., & De Arroyabe, J. C. F. (2024). Revealing the realities of cybercrime in small and medium enterprises: Understanding fear and taxonomic perspectives. *Computers & Security*, *141*, 103826. https://doi.org/10.1016/j.cose.2024.103826

Böhme, R., & Moore, T. (2012). The market for stolen data: Estimating the costs of data breaches. *IEEE Security & Privacy*, *10*(2), 43–49. https://doi.org/10.1109/MSP.2012.40

Boraine, P. A., & Leno Doris, D. N. (2019). The Fight against Cybercrime in Cameroon. *International Journal of Computer (IJC)*, *35*(1), 87–100. https://doi.org/10.53896/ijc.v35i1.1469

Chris, H. et al. (2005). *Criminology*. Oxford: Oxford University Press.

Edwards, M., & Hollely, N. M. (2023). Online sextortion: Characteristics of offences from a decade of community reporting. *Journal of Economic Criminology*, *2*, 100038. https://doi.org/10.1016/j.jeconc.2023.100038

Garner, B. A. (1999). *Black's Law Dictionary* (7th ed.). Minnesota, West Group.

Gupta, A., Matta, P., & Pant, B. (2025). Analyzing the social conduct of users to detect and prevent cybercrime across social networks. *Procedia Computer Science*, *258*, 4269–4278. https://doi.org/10.1016/j.procs.2025.04.676

Higgs, J., & Flowerday, S. (2025a). Detecting cybercrime in online video gaming. *Computers & Security*, *156*, 104528. https://doi.org/10.1016/j.cose.2025.104528

Higgs, J., & Flowerday, S. (2025b). Cybercrime and virtual laundering in video game marketplaces: A Bayesian P-inductive argument. *Technology in Society*, *83*, 102978. https://doi.org/10.1016/j.techsoc.2025.102978

Holt, T. J. (2022). Understanding the state of criminological scholarship on cybercrimes. *Computers in Human Behavior*, *139*, 107493. https://doi.org/10.1016/j.chb.2022.107493

Hornuf, L., Momtaz, P. P., Nam, R. J., & Yuan, Y. (2025). Cybercrime on the ethereum blockchain. *Journal of Banking & Finance*, *175*, *107419*. https://doi.org/10.1016/j.jbankfin.2025.107419

Jansson, K., & von Solms, R. (2011). Phishing for phishing awareness. *Behaviour & Information Technology*, *32*(6), 584–593. https://doi.org/10.1080/0144929x.2011.632650

Lee, S., Kang, I., & Kim, H. (2023). Understanding cybercrime from a criminal's perspective: Why and how suspects commit cybercrimes? *Technology in Society*, *75*, 102361. https://doi.org/10.1016/j.techsoc.2023.102361

Luu, T. J., Samuel, B. M., Jones, M., & Barnes, J. (2025). Exploring how the Dark Triad shapes cybercrime responses. *Personality and Individual Differences*, *244*, 113250. https://doi.org/10.1016/j.paid.2025.113250

Matias, C. F. F. (2025). Access revisited: AI training at the intersection of copyright and cybercrime laws. *Computer Law & Security Review*, *57*, 106149. https://doi.org/10.1016/j.clsr.2025.106149

Onwuadiamu, G. (2025). Cybercrime in Criminology; A Systematic Review of criminological theories, methods, and concepts. *Journal of Economic Criminology*, *8*, 100136. https://doi.org/10.1016/j.jeconc.2025.100136

Orbach, B., & Huang, L. (2018). Con men and their enablers: The anatomy of confidence games. *Social Research: An International Quarterly*, *85*(4), 795–822. https://doi.org/10.1353/sor.2018.0050

Payne, B. K. (2020). *The Handbook of International Cybercrime and Cyberdevidence*. London: Palgrave Macmillian. https://doi.org/10.1007/978-3-319-78440-3

Porcedda, M. G. (2023). Sentencing data-driven cybercrime. How data crime with cascading effects is tackled by UK courts. *Computer Law & Security Review*, *48*, 105793. https://doi.org/10.1016/j.clsr.2023.105793

Reyns, B. W., Henson, B., & Fisher, B. S. (2011). Being Pursued Online: Applying Cyberlifestyle–Routine Activities Theory to Cyberstalking Victimization. *Criminal Justice and Behavior*, *38*(11), 1149–1169. https://doi.org/10.1177/0093854811421448

Sarefo, S., Mphago, B., & Dawson, M. (2023). An analysis of Botswana's cybercrime legislation. *Procedia Computer Science*, *219*, 1023–1033. https://doi.org/10.1016/j.procs.2023.01.380

Sarkar, G., & Shukla, S. K. (2024). Reconceptualizing online offenses: A framework for distinguishing cybercrime, cyberattacks, and cyberterrorism in the Indian legal context. *Journal of Economic Criminology*, *4*, 100063. https://doi.org/10.1016/j.jeconc.2024.100063

Snidal, D. (2013). Rational Choice and Interntional Relations. In *Handbook of International Relations*. London, Sage. https://doi.org/10.4135/9781446247587.n4

Tok, Y. C., Zheng, D. Y., & Chattopadhyay, S. (2025). A Smart City Infrastructure ontology for threats, cybercrime, and digital forensic investigation. *Forensic Science International Digital Investigation*, *52*, 301883. https://doi.org/10.1016/j.fsidi.2025.301883

Wartiovaara, M. (2011). Rationality, REMM, and Individual Value Creation. *Journal of business Ethics*, *98*(4), 641–648. https://doi.org/10.1007/s10551-010-0643-6

Williamson, O. E. (2008). Outsourcing: Transaction Cost Economics and Supply Chain Management. *Journal of Supply Chain Management*, *44*, 2–82. https://doi.org/10.1111/j.1745-493x.2008.00051.x

Wright, D., & Kumar, R. (2023). Assessing the socio-economic impacts of cybercrime. *Societal Impacts*, *1*(1–2), 100013. https://doi.org/10.1016/j.socimp.2023.100013

Yi, M., Li, J., & Shen, G. (2024). Can targeted poverty alleviation reduce criminal offenses? Empirical evidence from China judgments online data. *China Economic Quarterly International*, *4*(1), 55–68. https://doi.org/10.1016/j.ceqi.2024.03.003

Yokotani, K., & Takano, M. (2021). Predicting cyber offenders and victims and their offense and damage time from routine chat times and online social network activities. *Computers in Human Behavior*, *128*, 107099. https://doi.org/10.1016/j.chb.2021.107099

Yokotani, K., & Takano, M. (2022). Effects of suspensions on offences and damage of suspended offenders and their peers on an online chat platform. *Telematics and Informatics*, *68*, 101776. https://doi.org/10.1016/j.tele.2022.101776

Сведения об авторе

Лекунзе Анумбуандем Бенволио – PhD, преподаватель, кафедра английского права, Университет Буэа
Адрес: Камерун, г. Буэа, а/я 63
E-mail: benleku@yahoo.com
ORCID ID: https://orcid.org/0009-0005-9947-0639

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс
Специальность ВАК: 5.1.4 / Уголовно-правовые науки

История статьи

Дата поступления – 27 апреля 2025 г.
Дата одобрения после рецензирования – 9 мая 2025 г.
Дата принятия к опубликованию – 25 сентября 2025 г.
Дата онлайн-размещения – 30 сентября 2025 г.