Criminal Liability in the Context of the Functioning of a Smart City

Wojciech Filipkowski <w.filipkowski@uwb.edu.pl>

Rafał Rejmaniak <r.rejmaniak@uwb.edu.pl>
BIAŁYSTOK, Poland

Abstract
The United Nations 2030 Agenda for Sustainable Development goals concern the functioning of individuals, societies, states, and the economy, and human interference with the environment. A smart city concept makes it possible to demonstrate and analyze many problems related to sustainable development in urban areas. This also applies to the use of artificial intelligence, which will manage such a city together with humans or without their active participation. All these issues are challenges for legal science. Moreover, the researchers (including lawyers) must not lose sight of the negative aspects of the actions taken by people and their organizations within that complex environment (including issues of criminal responsibility of humans).
The Authors rise general research questions: What is the concept of a smart city and what is the role of criminal law in this context? Their first goal is to present different roles of humans and AI concerning the functioning of smart cities as well as criminal acts that may be committed. There are e.g., end-users, manufacturers, developers, people responsible for implementing and maintaining services. The Authors present three concepts of attribution of criminal responsibility: decision loop, trustworthy artificial intelligence, and Human-Centered Automation. The criminal aspects related to the operation of completely autonomous AI systems are included in the consideration. Another goal is to present the solutions, reported in the doctrine, to evaluate human behavior in interaction with AI. The Authors discuss two major concepts: man-in-the-loop and man-on-the-loop.
Although these considerations are carried out in the context of the smart city, the Authors are convinced that the conclusions will be useful wherever such interaction is taking place or will take place in the future.

Keywords:
artificial intelligence, smart city, criminal responsibility

1. Introduction

The United Nations 2030 Agenda for Sustainable Development contains seventeen laudable goals that the international community would like to achieve over the next decade[1]. These include goals that concern the functioning of individuals, societies, states, and the economy, and human interference with the environment. Of key importance is proper understanding of the term "sustainable" in the context of social development, economic growth, and environmental development. By reference to the 1987 Report of the World Commission on Environment and Development entitled "Our Common Future," the term can be defined as improving the quality of life of people around the world without pillaging the earth's natural resources[2]. These are the two core priorities that are broken down into more specific priorities in the 2030 Agenda. They require differentiated action in different regions of the world in key areas, such as protection of natural resources and the environment, economic growth and equitable distribution of benefits, and human development. The key is to find the right balance between these priorities and areas.

A *smart city* concept, especially one including the environmental aspect of functioning of cities, makes it possible to demonstrate and analyze many problems related to sustainable development in urban areas. This also applies to the use of artificial intelligence in this area, which to a greater or lesser extent will "manage" such a city together with humans or without their active participation[3]. On the other hand, it is beyond dispute that conducting ongoing multidirectional activities (even by dynamically responding to the changes taking place) in order to improve the functioning of a city and the quality of life of its inhabitants will require the support of artificial intelligence.

---

1 About the Sustainable Development Goals - 'Take Action for the Sustainable Development Goals' <https://www.un.org/sustainabledevelopment/sustainable-development-goals> accessed on 14 April 2020.

2 World Commission on Environment and Development, 'Our Common Future. From one earth to one world' (Report, Annex A/RES/42/187, 11 December 1987).

3 United Nations, 'Artificial intelligence summit focuses on fighting hunger, climate crisis and transition to 'smart sustainable cities' (UN News, 28 May 2019) <https://news.un.org/en/story/2019/05/1039311> accessed on 14 April 2020.

The above idealistic, but also pragmatic, constructive, and progressive, view of the world and interstate relations, which underlies the 2030 Agenda, is a challenge for the legal science. However, while conducting research on a whole range of interrelated problems, one must not lose sight of the negative aspects of the actions taken by people and their organizations[4]. The question that arises is: What is the role of criminal law in this context? To answer this question, reference should be made to classical principles, such as subsidiarity and proportionality. The former indicates the role of criminal law as additional and complementary to the functioning of the legal system created by other branches of law[5]. It is used when other regulations have proved insufficient, but also to strengthen them. If one resorts to criminal law, its impact should be proportional to the criminal act committed[6]. Moreover, a detailed analysis must be made as to whether the application of criminal law norms has any undesirable side effects.

Taking into account the above assumptions, the goal that the authors set for themselves is to indicate the existing principles of criminal law in relation to cases of criminal acts committed by persons performing different roles in the system falling under the general term *smart city*. However, we do not lose sight of the fact that reality - and technology in particular - is changing, and the role of criminal law is also to respond to these changes. Therefore, another goal is to present the ways, reported in the doctrine, to evaluate human behavior in interaction with artificial intelligence. Although these considerations are carried out in the context of the *smart city*, we are convinced that the conclusions will be useful wherever such interaction is taking place or will take place in the future. On the other hand, we will not consider the question of criminal liability of artificial intelligence, because of our firm belief that this is premature[7].

---

4 Sławomir Redo, 'Chapter II. Priorytety Agendy na rzecz Zrównoważonego rozwoju 2030' in Emil Walenty Pływaczewski, Sławomir Redo, Ewa Monika Guzik-Makaruk, Katarzya Laskowska, Wojciech Filipkowski, Ewa Glińska, Emilia Jurgielewicz-Delegacz and Magdalena Perkowska, *Kryminologia, Stan i perspektywy rozwoju, Z uwzględnieniem założeń Agendy ONZ na rzecz zrównoważonego rozwoju 2030* (Wolters Kluwer 2019) 846-847.

5 Jan Kulesza, *Problemy teorii kryminalizacji. Studium z zakresu prawa karnego i konstytucyjnego* (Wydawnictwo Uniwersytetu Łódzkiego 2017) 156; Andrew Ashworth and Jeremy Holder, *Principles of Criminal Law* (7th edn., Oxford University Press 2013) 56.

6 Kulesza (n 5) 37-38; Ashworth and Holder (n 5) 33 (this principle is also referred to as "Criminalization as a last resort").

7 See: Ryan Abbott and Alex Sarch, 'Pushing Artificial Intelligence: Legal Fiction or Science Fiction' (2019) 53 University of California, Davis Law Review 332ff;

2. Assumptions of the smart city concept

2.1. Smart city and the 2030 Agenda

The term *smart city* was first used only in about 1992[8]. It seems, however, that the ranges of meanings presented by particular authors, or those contained in various types of documents issued to date, are different and have changed over the years. It depends on the authors' points of view, their education, the field of science they represent, and the goals set for these publications.

Many authors have made attempts to define the *smart city* concept[9]. One can try to put them on several levels. Smart cities are characterized by widespread presence of innovation processes that lead to creation of new products or improvement of existing products, technological processes, and organizational systems. Innovation as a process goes through a series of stages from idea to application to dissemination. Thus, this is a dynamic phenomenon that is gradually and unevenly implemented in the areas of functioning of a city (or many cities). Its course and intensity also depend on the degree of involvement of stakeholders, i.e. public authorities, inhabitants, private entities (municipal companies, businesses, start-ups), and non-governmental organizations. Another development factor is resources, primarily capital. On the other hand, an important feature of solutions that are being implemented is the aim to improve the quality of life of a city's inhabitants, but also of investors and tourists, through social and economic development. As we recall, these are the priorities of the 2030 Agenda.

Artificial intelligence - being an innovative tool in its own right - can be used to bring innovation to such areas as, for example[10]:

– management of the public sphere of the city - public management;
– offering products or services - e.g. public transport, technology parks;

Gabriel Hallevy, 'The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control' (2016) 4 Akron Intellectual Property Journal 177ff.

8 'Smart city. What is a smart city?' (Official website of the City of Vienna) <https://www.wien.gv.at/stadtentwicklung/studien/pdf/b008403j.pdf> accessed 14 April 2020.

9 Leonidas G. Anthopoulos, *Understanding Smart Cities: A Tool for Smart Government or an Industrial Trick?* (Springer 2018) 7-12; Alicja Korenik, *Smart cities, Inteligentne miasta w Europie i Azji*, (CeDeWu 2019) 19ff.

10 Korenik (n 9) 21.

- implementation and application of technological solutions - e.g. computerization of the city; and
- financial management - e.g. the provision of services in the form of a public-private partnership.

As the literature rightly points out[11], this concept is usually strongly identified only with implementation of a modern solution of a technological nature (especially IT). This is incorrect and is used by city authorities to serve marketing and promotional purposes. This can also happen with the use of artificial intelligence. The adjective *smart* should mean a greater ability to learn, collaborate, and above all solve problems of cities[12].

When describing the *smart city* concept in the context of the 2030 Agenda, it is impossible not to mention Goal 11, which is to make cities and human settlements inclusive, safe, resilient, and sustainable, and to involve all inhabitants in their functioning. This goal can be achieved by:

- ensuring access for all to adequate, safe, and affordable housing and basic services, and upgrading slums (11.1);
- providing access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport. Special attention should be paid to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons (11.2);
- enhancing inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries (11.3);
- strengthening efforts to protect and safeguard the world's cultural and natural heritage (11.4);
- significantly reducing the number of deaths and the number of people affected, and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, with a focus on protecting the poor and people in vulnerable situations (11.5);
- reducing the adverse per capita environmental impact of cities, including by paying special attention to air quality and municipal and other waste management (11.6);

---

11 ibid 24.

12 Waleed Ejaz and Alagan Anpalagan, *Internet of Things for Smart Cities* (Springer 2019)2ff; Łukasz Kowalski, 'Inteligentne miasta - przegląd rozwiązań' in Maria Soja and Andrzej Zborowski (eds), *Miasto w badaniach geografów* (Wydawnictwo Uniwersytetu Jagiellońskiego 2015) 105.

- providing easy and universal access to safe, inclusive, and accessible, green and public spaces, in particular for women and children, older persons and persons with disabilities (11.7);
  - supporting positive economic, social and environmental links between urban, suburban, and rural areas by strengthening national and regional development planning (11.a);
  - significantly increasing the number of cities and human settlements adopting and implementing integrated policies and plans towards inclusion, resource efficiency, mitigation and adaptation to climate change, resilience to disasters; and developing and implementing, in line with the Sendai Framework for Disaster Risk Reduction 2015-2030, holistic disaster risk management at all levels (11.b).

It can be assumed that artificial intelligence can make a significant contribution to achievement of these goals. This is due to the fact that having adequate computing power, access to data and information, as well as appropriate algorithms, it is able to forecast the effects of possible actions to be taken and to estimate to what extent these actions will achieve the objectives[13].

2.2. Smart areas

For a more complete analysis of the problem, it is also necessary to point out specific areas that allow evaluating the quality of services provided by cities and the quality of life in cities. In this respect, reference can be made to the set of ISO 37122:2019 standards (and earlier the ISO 3720:2014 standards), which are the result of cooperation between the European Union, the International Organization for Standardization in Geneva, and national standardization bodies. This is also an object of analyses carried out by experts[14]. Of the dozens of possible indicators, the following 17 are listed: education; fire and emergency response; safety; environment; economics; finance; recreation; health; telecommunications and innovation;

13 Thales, 'Secure, sustainable smart cities and the IoT' <https://www.gemalto.com/iot/inspired/smart-cities> accessed 27 February 2021.

14 Ejaz and Anpalagan (n 12) 2ff.

transportation; governance; energy; shelter; solid waste; water and sewers; wastewater; and urban planning[15].

All of them are important for raising the quality of life of city dwellers. What can be defined as the legal interests protected by criminal law are availability, integrity, and confidentiality in relation to a service in the broadest sense, updating the position represented in the doctrine of computer criminal law[16]. Firstly, it is in the interest of both the inhabitants themselves and the providers of services in these areas that they are available for use at any time and in the manner expected. Secondly, the integrity of a service (and in particular its associated data and information) is that no changes are made to it by unauthorized persons. They should be exactly as their suppliers assume and complete so as to perform some tasks effectively and keep them in proper condition. Thirdly, only authorized persons may have access to them. Modern information technology systems provide for different roles and associated ranges of authorizations to make changes to services, including, for example, entering, changing, or deleting data and information.

From the point of view of criminal law, four questions are relevant: To what extent were the above-mentioned interests violated or were they put at risk of indirect or direct infringement (state of danger)? What socially unacceptable consequences did the behavior cause? Who committed this act and what was his or her role in the system? How should the behavior be classified?

3. Selected technological issues

In the legal considerations being carried out, certain assumptions have to be made regarding technological issues:

– building a *smart city* will be done in a various ways for each area; at the moment individual cities around the world are implementing elements

15 'ISO 3720:2018. Sustainable development of communities — Indicators for city services and quality of life' (*International Organization for Standardization*, July 2017) <https://www.iso.org/standard/68498.html> accessed 14 April 2020.

16 Cf.: Andrzej Adamski, *Prawo karne komputerowe* (C. H. Beck 2000) 41-42.

of such systems[17], but so far they are choosing those that are most important to them[18];
- certainly, the very architecture of the system will have a modular nature: individual *smart* areas will ultimately constitute a functional whole, but at the same time they will be based on common data or will exchange data for more efficient functioning; individual modules will be upgradeable or exchangeable with others;
- the system will be distributed and network-centric;
- the implemented solutions will have different levels of automation and, in the future, different levels of artificial intelligence[19].

It seems that at the moment, the following technologies will be crucial for the functioning of current and future *smart cities*[20]:

- energy management on the scale from individual appliances, buildings, neighborhoods, up to the critical infrastructure of the city, which determines the use of any processes[21];
- transport management in the public spaces of cities, e.g. smart traffic lights, parking lots, and vehicles, all the way to synchronization of a multimodal transport system in the city and its surroundings[22];

17 Ejaz and Anpalagan (n 12) 11-14.

18 Richard van Hooijdonk, 'Top 10 smart cities that use tech to transform urban life' (Richard van Hooijdonk blog, 9 December 2019) <https://www.richardvanhooijdonk.com/blog/en/top-10-smart-cities-that-use-tech-to-transform-urban-life> accessed 14 April 2020.

19 Thomas B. Sheridan and Raja Parasuraman, 'Human-Automation Interaction' (2006) 1 Reviews of Human Factors and Ergonomics 89-129.

20 Cf. Mahashreveta Choudhary, 'Six technologies crucial for smart cities' (Geospatial world, 19 November 2019) <https://www.geospatialworld.net/blogs/six-technologies-crucial-for-smart-cities> accessed 14 April 2020; Teena Maddox, 'Smart cities: 6 essential technologies' (TechRepublic, 1 August 2016) <https://www.techrepublic.com/article/smart-cities-6-essential-technologies> accessed 14 April 2020./

21 George Koutitas, 'The Smart Grid: Anchor of the Smart City' in Stan McCleallan, Jesus A. Jimenez and George Koutitas (eds) *Smart Cities, Applications, Technologies, Standards, and Driving Factors* (Springer 2018) 53 ff.

22 Jesus A. Jimenez, 'Smart Transportation Systems' in Stan McCleallan, Jesus A. Jimenez and George Koutitas (eds) Smart *Cities, Applications, Technologies, Standards, and Driving Factors* (Springer 2018) 123 ff.

- efficient and secure (using e.g. *blockchain*) acquisition, collection, and analysis of large amounts of data (*big data*) and real-time decision-making (e.g. cloud computing[23]);
- *smart internet of things* - individual modules will have different levels of autonomy in their functioning in the system[24]; their basic functions are to acquire data and information (sensors[25]) or to put it into the system, transfer it to the elements dealing with collecting and further processing, as well as manipulators and switches, which make changes in the real world[26].

Each of the above technologies, in addition to its advantages, also leads to certain challenges, and in its extreme form also to dangers[27]. When used not in accordance their intended purpose, they can violate socially acceptable interests. This includes, for example, the life and health of their users; their freedom or privacy; the availability, integrity, and confidentiality of services, or the safety of individual users (or groups of users). It does not matter whether the perpetrator of these violations is public authorities, criminals, or other users of the elements that make up the entire *smart city* system. Besides, due to the interconnections between individual modules and areas, violations and threats in one part of the system can relatively easily spread to others and affect every person or organization operating in it. If only for this reason, research should be undertaken into the shaping of the criminal liability of the system participants.

---

23 Brad Booth, 'The Cloud: A Critical Smart City Asset' in Stan McCleallan, Jesus A. Jimenez and George Koutitas (eds) *Smart Cities, Applications, Technologies, Standards, and Driving Factors* (Springer 2018) 97 ff.

24 Ejaz and Anpalagan (n 12) 5 ff.

25 Soumia Bellaouar, Mohamed Guerroumi, Abdelouahid Derhan and Samira Moussaoui, 'Towards Heterogeneous Architectures of Hybrid Vehicular Sensor Networks for Smart Cities' in Zaigham Mahmood (ed) *Smart Cities, Development and Governance Frameworks* (Springer 2018)51 ff.

26 Raja Parasuraman, Thomas B. Sheridan and Christopher D. Wickens, 'A Model for Types and Levels of Human Interaction with Automation' (2000) 30 IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans 286-97.

27 Amrita Ghosal and Subir Halder, 'Chapter 5. Building Intelligent Systems for Smart Cities: Issues, Challenges and Approaches' in Zaigham Mahmood (ed) *Smart Cities, Development and Governance Frameworks* (Springer 2018) 119-120.

4. Basic problems of liability of smart city component users

4.1. Concepts of attribution of responsibility for a result to a human

Cooperation of a human with particular *smart city* modules can be shaped in different ways depending on the level of autonomy of the systems used in them: from treating such system as mere tools in the hands of a human, through cooperation of a human with the systems, to complete autonomy of the systems in making decisions and their implementation. Defining the role of humans and the tasks they are charged with is one of the fundamental challenges in the design and use of smart systems. One of the key issues in this regard is the problem of assigning responsibility - of whatever nature - for the effects caused by the operation of such systems. In the following sections, the decision loop, the concept of trustworthy artificial intelligence, and Human-Centered Automation will be presented.

4.1.1. Decision loop

The decision loop concept is used mainly to define the relationship between humans and the so-called systems with progressive autonomy of a military nature (*Lethal Autonomous Weapon Systems - LAWS*)[28], but it can also be successfully applied to *smart city* components. It distinguishes three models of the relationship between humans and systems. The first model, called man-in-the-loop, describes a situation in which systems have no freedom of action. Humans, on the other hand, control the systems and make decisions[29] or authorize them before they are implemented[30]. Humans are directly responsible for the operation of the systems. The systems themselves are a tool in the hands of the operators, like a hammer or a screwdriver. An example of this type of cooperation between humans and systems is the use of a remotely controlled drone by its operator.

---

28 Jeffrey J. Caton, *Autonomous Weapon Systems: A Brief Survey of Developmental, Operational, Legal, and Ethical Issues* (United States Army War College Press 2015) 3-4; Ajey Lele, 'Debating Lethal Autonomous Weapons Systems' (2019) 13 Journal of Defense Studies 55-56; William C. Marra and Sonia K. McNeil, 'Understanding "The Loop": Regulating the Next Generation of War Machines' (2012) 36 Harvard Journal of Law & Public Policy 1141-1142.

29 Noel Sharkey, 'Saying 'No!' to Lethal Autonomous Targeting' (2010) 9 Journal of Military Ethics 370.

30 Lele (n 28) 55-56.

The second model describes a situation where systems act autonomously: they make and implement "decisions" without an active role played by humans. Humans do not take part in the decision-making process, but they supervise the operation of the system and can interrupt it ("veto") when they consider it necessary. In this model, referred to as man-on-the-loop, humans are responsible for preventing the systems from implementing wrong decisions.

This concept is further distinguished by a third model, referred to as man-out-of-the-loop. In this model, the system operates completely autonomously with no human oversight. The role of humans is merely to give the system commands or define its tasks. However, humans do not have the power to stop its operation. This model is controversial as it raises significant problems in terms of the possibility of attributing responsibility for the "actions" of the systems to humans. Humans do not know how the systems will perform the set tasks, and have no way to react when unforeseen circumstances arise or when the systems behave differently than in the assumed scenario. Advocates of full autonomy of the systems try to create algorithms to ensure proper operation of the systems in unforeseen circumstances[31].

In characterizing this concept, it should be added that the systems may exhibit different levels of autonomy in their performance of different tasks. The decision-making process can be decomposed into smaller parts, e.g., based on Boyd's loop ("OODA Loop")[32], which divides the decision-making process into 4 components: observe (data collection), orient (analysis), decide, and act. At the different stages of the decision-making process, the system may have varying degrees of autonomy. For example, it can be completely autonomous in the observe and orient stages, and supervised by a human in the decide and act stages[33].

---

31 Ronald C. Arkin, Patrick Ulam and Brittany Duncan, 'An Ethical Governor for Constraining Lethal Action in an Autonomous System' (Technical Report GIT-GVU-09-02, Georgia Institute of Technology Atlanta Mobile Robot Lab, 2009).

32 A diagram of the Boyd's loop is available at: Boyd JR, 'The Essence of Winning and Losing' (September 2012) <https://fasttransients.files.wordpress.com/2010/03/essence_of_winning_losing.pdf> accessed on 14 April 2020.

33 Marraand McNeil (n 28) 1146-1147.

4.1.2. Trustworthy artificial intelligence

As for determination of the appropriate entity and the grounds for attributing criminal liability to it for the negative effects on legal interests caused by the operation of *smart city* components, it should be emphasized once again that these components usually use artificial intelligence systems to a greater or lesser extent. It is the characteristics of artificial intelligence systems, which may be their advantages, that also cause significant difficulties in determining the entity responsible for their actions. These systems have the capacity to learn and exhibit autonomy from the human interacting with them[34]. Furthermore, it is sometimes impossible to determine why an AI system made a particular decision[35].

Recognition of these problems has led the European Union to formulate and practically develop the concept of "trustworthy artificial intelligence". Seven requirements have been formulated as a part of the development of basic European standards for the design and use of AI systems[36]:

- human agency and oversight;
- technical robustness and safety;
- privacy and data governance;
- transparency;
- diversity, non-discrimination, and fairness;
- social and environmental wellbeing; and
- accountability.

With respect to conditions directly related to accountability for the actions of such systems, the European Commission indicated that one of the key requirements is to ensure an appropriate degree of control measures depending on the specific AI system and its area of application[37]. In addition, such systems should be assessed by both internal and external auditors,

34 See: Tomasz Zalewski, 'Definicja sztucznej inteligencji' in Luigi Lai and Marek Świerczyński (eds) *Prawo sztucznej inteligencji* (C. H. Beck 2020) 11.

35 See for example: Anna Kasperska, 'Problemy zastosowania sztucznych sieci neuronalnych w praktyce prawniczej' (2017) 11 Przegląd Prawa Publicznego 25.

36 Commission, 'White Paper on Artificial Intelligence. A European approach to excellence and trust', COM (2020) 65 final 11.

37 Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Building Trust in Human Centric Artificial Intelligence', COM (2019) 168 final 5.

and any potential adverse effects of the use of such systems should be identified, assessed, and documented[38].

The High Level Expert Group on Artificial Intelligence established by the European Commission in 2018, with respect to the condition of ensuring a oversight human role, indicated that such oversight can take place in different ways, according to the principle of human participation (*human-in-the-loop* (HITL) - assuming the possibility of human intervention in each decision cycle of the system), the principle of human intervention (*human-on-the-loop* (HOTL) - assuming the possibility of human intervention during the system design cycle and monitoring of the system operation), or the principle of human control (*human-in-command* (HIC) - the possibility of only supervising the general functioning of the system and deciding when and how the system will be used)[39]. The need to meet the condition of ensuring an oversight role for humans was also indicated in the Policy for the development of artificial intelligence in Poland until 2020[40].

4.1.3. Human-Centered Automation

Another concept that should be pointed at is *Human-Centered Automation* (HCA). Its basis is the assumption that systems must be designed to cooperate with humans or to interact with them in other ways[41]. According to this concept, humans are responsible for proper functioning of systems. Consequently, systems should be designed in such a way that the human operator is in command in this collaboration (although the implementation of this assumption in individual cases may cause some difficulties[42]). This is possible when the operator is involved in the operation of the sys-

38 ibid 7.

39 High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy Artificial Intelligence' (2019) 20.

40 Resolution no. 196 of the Council of Ministers of 28 December 2020 on establishing the "Policy for the development of artificial intelligence in Poland until 2020," Monitor Polski 2021, item 23, Annex, 66.

41 Sheridan and Parasuraman (n 19) 94.

42 Toshiyuki Inagaki, 'Design of human-machine interactions in light of domain dependence of human centered automation' (2006) 8 Cognition, Technology & Work 164.

tem and is duly informed of the principles of its operation[43]. This concept also requires that the operation of the system is predictable and that the human ability to monitor it is real[44]. In addition, the system should also monitor human behavior and both the system and its operator should communicate their "intentions" in the task undertaken (what they intend to achieve)[45].

4.2. The question of full autonomy of systems

In addition to the concepts of human-system interaction presented above, other stances can be found. It seems reasonable to mention the increasingly clear view that the most desirable state is full autonomy of systems, especially with regard to artificial intelligence. Although it is difficult to assume nowadays that full autonomy could characterize cooperation of humans with various systems in general, this assumption can be developed in some fields.

The idea of striving for a fully autonomous system guides, for example, developers of the concept of autonomous vehicles that drive, sail, or fly[46]. Developing this concept may be important for the functioning of *smart cities*[47]. Dissemination of such vehicles will change the urban landscape (less space will be used for parking), reduce emissions, decrease congestion, and the vehicles themselves can also be used for public transport[48]. At the same time, the problem of legal regulation of the functioning of such solutions is becoming more and more urgent. The legal solutions obtained

43 Charles E. Billings, 'Human-Centered Aviation Automation: Principles and Guidelines' (NASA Technical Memorandum 110381, Ames Research Center, February1996) 8 ff.

44 See for example: Cheng Zhang and others, 'Human-centered automation for resilient nuclear power plant outage control' (2017) 82 Automation in Construction 182.

45 Billings (n 43) 11 ff.

46 Cf.: Tomasz Neumann, 'Perspektywy wykorzystania pojazdów autonomicznych w transporcie drogowym w Polsce' (2018) 19 Autobusy 787-788; 'Dronomat zamiast paczkomatu. Czy wkrótce niebo zaroi się od dronów?' (*Rozmowy Instytutu Nowej Europy* on Anchor.fm, 10 June 2020) <https://anchor.fm/instytutnowejeuropy/episodes/Dronomat-zamiast-paczkomatu--Czy-wkrtce-niebo-zaroi-si-od-dronw-ef77ru> accessed 27 February 2021.

47 Testing of autonomous vehicles is one of the indicators for evaluating smart cities in the *Global Smart City Performance Index*. See: Korenik (n 9) 32.

48 Neuman (n 46) 789-790.

in this way can provide experience and possibly be a starting point for the development of further standards for other systems of this type.

However, it should be emphasized that the concept of full autonomy currently faces a number of difficult and as yet unresolved issues concerning the question of attribution of responsibility for the effects caused by such systems[49]. In a situation where there is no human being who could bear responsibility for the damage caused by such a system, the following questions arise: Can the system itself be a subject capable of bearing responsibility? On what principles would it be liable? Would it be entitled to any guarantees and rights? What sanctions could be imposed on it? Although the discussion regarding these problems has already begun, it seems premature to consider the possibility of attributing legal liability (including criminal liability) to the system itself in the current conditions[50]. However, the issue of manufacturers or suppliers of such solutions and their liability remains open.

4.3. Scope of responsibilities of system users

The entity involved in performance of the tasks of a smart system is its user. The user can be defined as the entity that has purchased and implemented, and is using the system (a legal person, e.g. a company under private law or local government bodies). Typically, the functioning of such a system is further based on its cooperation with or supervision by a human (operator). Another category of system users is people who only use products or services that have already been implemented, e.g. residents of *smart cities*. In the following discussion, we will analyze the first example, because people act on behalf of the legal entity that implements a system.

The responsibilities of an entity that uses an information-technology system can be divided into responsibility to ensure proper operation of the system itself, responsibility to ensure competence of the operators and supervisors of the system, and responsibility to establish appropriate security procedures in case of a threat. The first category of responsibilities includes ensuring proper control and authorization of access to the system,

---

49 See for example: Sabine Gless, Emily Silverman and Thomas Weigend, 'If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability' (2016) 19 New Criminal Law Review 412–436.

50 Cf. Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Springer 2015) 229; Woodrow Barfield and Ugo Pagallo, *Advanced Introduction to Law and Artificial Intelligence* (Edward Elgar Publishing 2020) 120-121.

use of encryption (e.g. *blockchain*), regular security audits by an external entity[51], and making sure to update the system and maintain it.

The second category of responsibilities on the part of the entity that uses a system is to ensure that the people who work with or oversee it are adequately prepared. It is therefore necessary that these persons are properly trained and have sufficient skills to carry out the tasks they are charged with. This is the necessary condition of real, and not only formal, human participation in the decision-making process carried out by the system. The literature on human interaction with smart systems analyzes many undesirable phenomena that need to be addressed. These include:

– excessive confidence in the operation of the system (*overreliance*)[52];
– lack of situation awareness when a human oversees the system[53], and
– loss of the skills needed when taking control of the system's tasks (*deskilling*)[54].

The entity that implements a smart system should also define the internal procedures to deal with malfunctions that[55] should be communicated to those working with it.

Another important responsibility of the entity that using a smart system is to collect and secure logs. In a situation where the operation of the system has caused damage, analysis of previously stored data may make it possible to determine the cause of the event, to detect anomalies in the operation of the system, and to trace and evaluate the actions taken by the operator[56].

In contrast, the responsibilities of a human interacting with an smart system may take different forms depending on the characteristics of the system in question. If the system shows a low degree of autonomy, e.g. it only suggests taking an action on the basis of analyzed data, the human

---

51 Indu B. Singh and Joseph N. Pelton, 'The cyber city of the future' (2013) 47 The Futurist 23.

52 Sheridan and Parasuraman (n 19) 98-100.

53 Mica R. Endsley, 'Automation and situation awareness' in Raja Parasuraman and Mustapha Mouloua (eds) *Automation and Human Performance. Theory and Applications* (reprint, CRC Press 2009) 178.

54 John D. Lee and Bobbie D. Seppelt, 'Human factors and ergonomics in automation design' in Gavriel Salvendy (ed) *Handbook of Human Factors and Ergonomics* (John Wiley & Sons 2012) 1616, 1617.

55 Zubair A. Baig and others, 'Future challenges for smart cities: Cyber-security and digital forensics' (2017) 22 Digital Investigation 7.

56 Ben Shneiderman, 'Human Responsibility for Autonomous Agents' (2007) 22 IEEE Intelligent Systems 61.

make the decisions, while the system itself should be treated as a mere tool used to perform its tasks. Assuming that the system works properly and collects the appropriate data, it will act as an "adviser", but making the final decision and taking the right action (or giving instructions for such action) will be the responsibility of the operator[57]. It is therefore important for the human to be aware of the role of the smart system as a decision support and its limitations (e.g., in terms of the tasks for which it was designed or the probability of the forecasts it formulates). Otherwise, the operator who formally decides to take certain actions will in fact be just a "human stamp" authorizing the decisions of the system, without the possibility of their assessment. However, the use of an advisory system may give rise to the temptation for its users to treat the system as a decision-making authority. Such a system would thus provide a mental "moral buffer" against responsibility for making difficult decisions, giving the illusion that it is the system, and not the person, that is responsible for the consequences of the choice made[58]. Depending on the specific nature of the operator's tasks and the purpose of the system used[59], the content of his or her duties should be defined in an agreement with his or her employer.

The situation is slightly different when the system has greater autonomy. The human then plays the role of a supervisor and his or her primary task is to monitor whether the system is functioning properly. The responsibilities of such a person should therefore include interrupting the system when it performs its tasks incorrectly (*mitigating the failure*). Depending on the specific characteristics of the system, this may involve, for example, an obligation to turn off the system, stop the implementation of the decision made by the system (*veto*)[60], or take control of the task and implement it "manually" (in systems that use *adaptable automation*[61]). Again, however, it is important to note that the supervisor must be properly trained and competent. Lack of sufficient knowledge and experience can lead to situations where the supervisor fails to notice an existing problem in the performance of the system, although he or she should notice it (*omission error*), or, without due verification, considers the operation of the

57 Mary L. Cummings, 'Automation and Accountability in Decision Support System Interface Design' (2006) 32 The Journal of Technology Studies 28.
58 ibid 26.
59 Ben Wagner, 'Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems' (2019) 11 Policy & Internet 115, 117.
60 Shneiderman (n 56) 60-61.
61 Lee and Seppelt (n 54) 1626-1627.

system to be correct when, in light of other existing data, it would appear to be faulty (*commission error*)[62].

5. *Selected specific problems of criminal liability of users of smart city components*

The issue of assignment of criminal liability for exposure or infringement of legal interests caused by smart systems is a challenge for the science of criminal law. The tendency to blame the operator for any malfunctions of the system also causes difficulties[63]. When examining the issue of criminal liability of a system operator or supervisor (understood as an individual) for the operation of the system, two models of interaction between humans and smart systems must be distinguished. Using the decision loop concept, these can be referred to as *man-in-the-loop* and *man-on-the-loop*.

When analyzing the principles of criminal liability of the user of a smart system for the effects caused by that system, it should be noted that liability incurred by a human operator or supervisor is only one possibility. There is also the possibility of civil liability of both individuals and legal persons. However, this issue is beyond the scope of this paper.

5.1. Man-in-the-loop

In the first case, it is the human who is in control of the system and who makes the decisions, so he or she is directly responsible for the effects caused by the decisions. An offence committed in such conditions is generally be a crime of commission. The effect of exposing or violating a legal interest to danger is the result of the perpetrator's active behavior. Depending on the specific situation, the user's criminal liability may take different forms.

If the operator consciously made a decision, knowing that its execution would cause the damage (he or she knew it and wanted to commit a criminal act, or although he or she did not want to commit it, he or she

62 Linda J. Skitka, Kathleen Mosier and Mark D. Burdick, 'Accountability and automation bias' (2000) 52 International Journal of Human-Computer Studies 701-702.

63 Karen Hao, 'When algorithms mess up, the nearest human gets the blame' (MIT Technology Review, 28 May 2019) <https://www.technologyreview.com/2019/05/28/65748/ai-algorithms-liability-human-blame> accessed on 26 April 2020.

accepted it), he or she committed an intentional act and used the system as an instrument of crime. Examples of such offences are murder, causing loss of health, causing a disaster, or putting in danger.

Much more difficulties are caused by a situation in which the user acted unintentionally, i.e. when he or she anticipated the possibility of committing a prohibited act and groundlessly believed that he or she would manage to avoid it, or did not anticipate such a possibility at all, although he or she could have and should have foreseen it (provided that the legislator allows for the possibility of incurring criminal liability for an unintentional offence). The user, by contract (or certain regulations), assumes the obligation to perform certain tasks (*task responsibility*)[64]. If, nevertheless, the constitutive elements of the offence are present and a cause and effect relationship has been established between the operator's conduct and the result, it is necessary to examine whether the operator complied with the precautionary rules required in these circumstances. They may be defined by e.g. the rules of system use defined by its supplier or internal procedures to be followed when carrying a given task. An example is the requirement to verify the system's suggestions on the basis of independent data instead of unreflective acceptance of the recommended solutions[65].

Determination of criminal liability for unintentional crimes, however, involves significant challenges that need to be met. The first is the possibility of assigning criminal responsibility to the operator when the required response time in a dynamic environment has exceeds the biological capabilities of a human being. Various smart systems can perform tasks at different speeds. According to the criterion of the system response time, tasks can be divided into strategic - in which the time of task completion is specified in minutes-days, tactical - in which the response time is from 5 seconds to several minutes, and operational - which are performed in less than 5 seconds[66]. Systems (primarily those using artificial intelligence) that perform the tasks assigned to them in real time may require extremely fast human response. On the one hand, it is the task of the system supplier to design a system suitable for the human perceptual capabilities. On the other hand, it is impossible to predict all situations in the dynamic environment in which the system will be used. For example, it may be

64 Giuseppe Contissa, 'Automation and Liability: an Analysis in the Context of Socio-Technical Systems' (2017) 11 i-lex 20-21.

65 ibid 23.

66 Lee and. Seppelt (n 54) 1625.

necessary to suspend instructions given to the system due to an unforeseen change in the situation that requires a change in the operator's decision. It seems that in such cases it should be examined whether the operator was objectively able to avoid the effect, because if this was not the case, the operator is not liable under criminal law, because one cannot require him or her to do something that is impossible.

Further problems may arise from multiple operators and smart systems working in parallel. It may happen that an isolated decision of a particular operator does not in itself turn out to be wrong. Instead, only in a specific situational setting created by a simultaneous operation of multiple systems and operators its fallacy becomes apparent. System malfunctions can be the result of the sum of independent errors made by different operators or people on different levels of the organizational hierarchy. In other words - a theoretically correct decision made in a certain factual situation can cause a damage resulting from the existence of a network of interdependencies between different elements of the system (systems) that created at a given moment an arrangement that is conducive to the occurrence of damage ("*a many hands problem*")[67].

5.2. Man-on-the-loop

The second category of cases are situations where the system has a higher degree of autonomy and makes decisions on its own, while the human monitors the performance of the task. Thus, the user assumes the position of a supervisor, passively observing the system, and is obliged to interrupt its operation in case of a threat (and possibly to take control manually).

Where the operation of the system puts in danger or violates interests protected by criminal law, the supervisor may be accused of failing to stop the operation when he or she could and should have done so. Failure to fulfill the duties (breach of precautionary rules) imposed on the supervisor constitutes grounds for attributing criminal liability to him or her. The human is thus the guarantor of non-occurrence of the effect. Under criminal law, a guarantor is a person who is under a specific legal obligation to prevent an effect that is a constitutive element of a given type of crime. The specific nature of the obligation means that its addressee is not everyone, but only those who have certain characteristics that distinguish them

67 Contissa (n 64) 29.

due to the relation to the interest protected by a legal norm[68]. The legal nature, on the other hand, indicates that the obligation must be grounded in law[69]. Although there is a dispute in the doctrine of criminal law as to what may be the source of such a duty, the catalog of such sources indicates a statute and a voluntary acceptance of a duty to prevent an effect (e.g. an employment contract)[70].

If the supervisor allows a damage caused by the smart system to occur by failing to stop its operation, he or she may be criminally liable for a crime of omission. The accusation against the supervisor, however, does not concern the fact that he or she caused the effect, but the fact that he or she did not take the necessary steps to prevent such an effect, although he or she was legally obliged to do so. It should be emphasized that such liability could be incurred by the supervisor only if the effect[71] could have been objectively foreseen and prevented[72]. This is because a guarantor cannot be required to do something that cannot be done[73]. Otherwise the form of his or her responsibility would be dangerously close to the construction of objective responsibility, independent of the existence of fault, which has been gradually abandoned since the Middle Ages[74]. Liability for culpable acts, on the other hand, is the foundation of modern criminal law, expressed synthetically in the *nullum crimen sine culpa* principle[75].

---

68 Maciej Kliś, 'Źródła obowiązku gwaranta w polskim prawie karnym' (1999) 2 Czasopismo Prawa Karnego i Nauk Penalnych 173.

69 Alicja Grześkowiak, 'Komentarz do art. 2 k.k.' in Alicja Grześkowiak, and Krzysztof Wiak (eds.), *Kodeks karny. Komentarz* (6th edn, C. H. Beck 2018) 50-51;Kliś (n 68) 170.

70 Grześkowiak (n 69) 51.

71 Jacek Giezek, 'Teorie związku przyczynowego oraz koncepcje obiektywnego przypisania' in Ryszard Dębski (ed.), *System prawa karnego, tom 3: Nauka o przestępstwie. Zasady odpowiedzialności* (C. H. Beck 2017) 547-548.

72 Andrzej Zoll, 'Komentarz do art. 2 k.k.' in Włodzimierz Wróbel and Andrzej Zoll (eds) K*odeks karny. Część ogólna. Tom I. Komentarz do art. 1-52* (Wolters Kluwer 2016) 89; Damian Tokarczyk, 'Obowiązek gwaranta w prawie karnym' (2014) 76 Ruch Prawniczy, Ekonomiczny i Socjologiczny 211.

73 Tokarczyk (n 72) 208.

74 Robert Zawłocki, 'Pojęcie przestępstwa' in Ryszard Dębski (ed.), *System prawa karnego, tom 3: Nauka o przestępstwie. Zasady odpowiedzialności* (CH Beck 2017) 52; Wacław Urszczak, *Historia Państwa i Prawa Polskiego. Tom I (966-1795)* (Wolters Kluwer 2013) 115.

75 Andrzej Zoll, 'Komentarz do art. 1 k.k.' in Włodzimierz Wróbel and Andrzej Zoll (eds.), *Kodeks karny. Część ogólna. Tom I. Komentarz do art. 1-52* (Wolters Kluwer 2016) 76.

It should be emphasized that the guarantor's liability does not exclude the possibility that the manufacturer of the system is also liable for the effect caused. However, determination of whether, in fact, the circumstances in which the system made a faulty decision could have been objectively foreseen at the system design stage continues to be a challenge[76].

The legal consequences of a system supervisor's omission may vary depending on whether the guarantor took any measures to counteract the effect and whether these were adequate to eliminate the danger, what his or her intent was, what effect materialized, whether it is possible to assess from a hindsight perspective how the danger would have been affected if he or she had taken appropriate action[77].

However, if the supervisor takes over the tasks of the system to perform them manually, the he or she assumes the role of a direct operator (*man-in-the-loop*), with all the consequences this entails.

6. Conclusion

As has been shown, the basic issue that determines the scope of criminal liability of an individual is the specific design of the system in which human interaction with artificial intelligence takes place. In the case of the *smart city* concept, we assumed that the system will have a modular structure and its individual components (services, products) will be at different levels of implementation of automated solutions up to those using artificial intelligence. This assumption results from an observation and analysis of the practice of implementation of this concept to date. The human role in each module may vary depending on the level of autonomy of the systems used in them.

This influences the specification of the duties of a human operator or supervisor of such smart modules, subsystems, services, products, etc. At the same time, it is a challenge for both the designers of individual system components (e.g. by defining the principles of task allocation between the human and the system and of conflict resolution - if any conflicts are allowed) and the entity that implements and uses them (e.g. ensuring proper system configuration, training for operators, and counteracting negative

76 Wojciech Filipkowski, 'Prawo karne wobec sztucznej inteligencji' in Luigi Lai and Marek Świerczyński (eds) *Prawo sztucznej inteligencji* (C. H. Beck 2020) 124-125.

77 Tokarczyk (n 72) 211-212.

phenomena occurring during cooperation between the human and the system).

So far, various concepts have been developed to define the principles of human-system interaction. Some of them require constant present of a human in the decision-making process of the system and the human's responsibility for the system's operation (Trustworthy Artificial Intelligence, *Human-Centered Automation*). There are also proposals to develop full autonomy of systems. *Human-out-of-the-loop* is one of the models in the decision loop concept; this direction is also adopted in the design of autonomous vehicles. From the standpoint of criminal law, which is based on human responsibility for prohibited acts and requires the presence of guilt (a person is accused the fact that in the situation in which he or she found himself, he or she could and should have behaved differently than he or she did), allowing fully autonomous systems to function would generate hitherto unresolved significant problems in determining the subject to whom responsibility should be attributed if the constitutive elements of a crime are in place.

For systems with a low degree of autonomy (falling within the *man-in-the-loop* model), the operator is responsible for the effects they cause. In such a situation, a systems is used as a tool (e.g. an "adviser" in decision making or an "executor" of a command given by a human). When humans cooperate with systems with higher levels of autonomy, humans assume the role of supervisors. The concept of a guarantor present in criminal law can be successfully used to determine the principles of a human's liability. This concept makes it possible to accuse the supervisor of failing to prevent the occurrence of an effect in a certain situation when he or she was legally obliged to do so (the so-called guarantor of non-occurrence of an effect).

However, there are some challenges associated with the issue of criminal liability of users of smart systems, such as those related to the cooperation and interdependence of different systems ("*many hands problem*") and to biological limitations of humans. In addition, there is the problem of examination of the level of awareness of the operator or the supervisor of the smart system, his or her knowledge of the procedures, and the principles on which the solution that makes up the *smart city* concept is based.

In conclusion, it should be emphasized that it is necessary to conduct research in this area of criminal law. Moreover, this research must be interdisciplinary. Lack of expert knowledge from different areas makes it difficult to establish communication between researchers, but can also lead to ill-considered and unforeseen consequences. However, it is beyond

dispute that such research is necessary if we want to achieve Goal 11 of the 2030 Agenda, which is to make cities and human settlements inclusive, safe, resilient, and sustainable, and to involve all inhabitants in their functioning with the use of artificial intelligence.