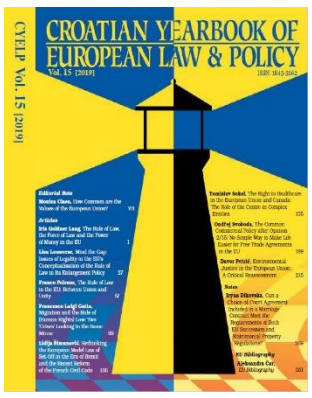

Department of European Public Law
Faculty of Law – University of Zagreb

Croatian Yearbook of European Law and Policy

ISSN 1848-9958 (Online) | ISSN 1845-5662 (Print)

Journal webpage: https://www.cyelp.com

Is it Time to Consider EU Criminal Law Rules on Robotics?

Igor Vuletić and Tunjica Petrašević

Suggested citation: I Vuletić and T Petrašević, 'Is it Time to Consider EU Criminal Law Rules on Robotics?' (forthcoming in 2020) 16 CYELP.

Link: https://www.cyelp.com/index.php/cyelp/article/view/382

IS IT TIME TO CONSIDER EU CRIMINAL LAW RULES ON ROBOTICS?*

Igor Vuletić** and Tunjica Petrašević***

Abstract: This paper is devoted to issues which have not been sufficiently explored in European literature, and which have had fragmented consideration in comparative literature. These issues raise the question of whether the EU legislator should develop a framework of criminal law rules which would regulate the use of Artificial Intelligence (hereinafter: AI) in the near future, and what such rules should specifically address. The authors recognise two issues of particular importance for the future regulation of AI development within the EU, and offer their perspective on the areas which should be subjected to regulation in this regard. In order to provide a systematic overview of this topic, the paper starts with a description of the recent regulatory action of the EU in the field of AI, with special reflection on the Ethics Guidelines for Trustworthy AI. The authors then describe what are, in their opinion, the most important intersections of AI and criminal law in the broader sense, and in conclusion present their views of which areas should be specifically regulated by EU legislature in this context.

Keywords: robot, Artificial Intelligence, criminal law, criminal procedure, autonomous, sanctions, European Union.

1 Introduction

The term 'robot' was first introduced by Czech author Karel Čapek in his renowned SF drama *Rossumovi Univerzální Roboti* of 1920.[1] This term subsequently became the colloquial name for all types of devices with a certain level of independence in their operation. For the purposes of this paper, the definition of robot as provided by Jack Balkin will be used. According to Balkin, the term robot applies to 'all material objects that interact with their

* The research for this paper was partly conducted within 'Artificial Intelligence and Criminal Law (IP-PRAVOS-18)', a project funded by the Faculty of Law Osijek.

** Associate Professor of Criminal Law at J. J. Strossmayer University in Osijek.

*** Associate Professor of Constitutional and EU Law at J. J. Strossmayer University in Osijek.

[1] See Karel Capek and Ivan Klima, *R. U. R.* (Claudia Novack-Jones (tr), Rossum's Universal Robots edition 2004).

environment, artificial intelligence agents, and machine learning algorithms'.[2] This definition is adequate because it also covers other forms of artificial intelligence which do not have physical embodiments in space (such as software).[3] It is clear that the terms 'robot' and 'AI' are not synonymous; certain authors have taken the position that AI is a broader, generic term, because robots can function by means of artificial intelligence, but not necessarily, because a robot can be a mechanical device, ie an extension of a human arm.[4] The position of the European Commission speaks in favour of the position that: 'AI-based systems can be purely software-based, acting in the virtual world (eg voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (eg advanced robots, autonomous cars, drones or Internet of Things applications)'.[5] The European Parliament also offered a definition of so-called 'smart robots' which implies robots which function by means of artificial intelligence with a degree of autonomy.[6] For the purposes of this paper, we are only interested in such 'smart robots' which are based on artificial intelligence. The terms 'robot' and 'artificial intelligence' or 'AI' will be used here interchangeably as synonyms in this paper.[7]

Devices functioning on AI can also cause significant damage to persons and property, which will be described in more detail later in this paper.

The development of artificial intelligence is a trend which creates great opportunities for progress and for enhancing the quality of life. There is even talk of a fourth industrial revolution or 'Industry 4.0', a phrase coined by the German government and which has enjoyed wide reception.[8] The infiltration of AI into various spheres of human life has numerous advantages, because it improves the quality of the provision of services in various sectors to the point that there

---

[2] Jack M Balkin, '2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data Lecture' (2019) 78 Ohio State Law Journal 1217, 1219.

[3] On the other hand, there are authors who link the term 'robot' exclusively to the types of AI which have a certain embodiment in the outside world. See, for example, Ryan Calo 'Robots in American Law' (2016) University of Washington School of Law Research Paper 2016-04, 6, available at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737598> accessed 20 March 2020.

[4] See Sunčana Roksandić Vidlička, Lelde Elīna Liepiņa and Svitozar Ostapchuk, 'Bioethical and Legal Challenges of Artificial Intelligence and Human Dignity' in Miodra Jovanović and Tibor Virady (eds), *Human Rights in the 21st Century* (Eleven International Publishing 2020) 273.

[5] Commission, 'Artificial intelligence for Europe' (Communication) COM(2018) 237 final, 1.

[6] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) para 1.

[7] Here it must be noted that the definition of the term 'robot' can be approached from various angles. However, this will not be elaborated further, as it would go beyond the scope of this paper. For more on this, see, for example, Thomas Kirchberger, 'European Union Policy-Making on Robotics and Artificial Intelligence: Selected Issues' (2017) 13 Croatian Yearbook of European Law and Policy 191, 196-197.

[8] ibid, 192.

is even valid apprehension that robots might replace the human workforce in most occupations in the near future.[9] This trend raises many concerns, among which the legal ones are the most significant. The legal significance of this phenomenon is best reflected in the fact that in some European countries a separate legal branch – the laws of robotics – is already being developed, and that it is being examined at academic, legislative and practical levels.[10]

One of the questions which law has to answer is how to regulate the operation of AI in such a manner as to ensure legal certainty but in a way which will not hinder the development of modern technology and the progress of society. This topic can be approached from various angles: labour law, civil law, commercial law, consumer protection law, etc. However, the concern here is the analytical approach to this topic from the perspective of criminal law. It is already evident that robots are taking over the roles of humans, and not only in ways which are useful for society, but also in harmful and unwanted ways.

If robots are imposed in addition to or as an alternative for humans in various occupations where there is a potential for damage, then there is clearly a possibility that such damage can occur when such occupations are performed by robots.

If a human driver can cause a traffic accident, a robot can do so as well. What remains unclear at this point is: who bears criminal liability in the latter case? This issue is not limited only to the field of traffic safety, but also applies to other fields where autonomous robotic technologies are being introduced (the weapons industry, medicine, financial operations, etc). Considering that the EU internal market follows global trends and is on track with the adoption of the modern technological trends of our time, there is no doubt that this issue is quite relevant in the context of EU law. This paper will therefore answer two questions which, in our view, have been disregarded up to now: are there criminal law fields which are ripe for regulation (and which specific fields are these), and is the EU the appropriate level for the regulation of these fields (and what exactly should be regulated) and does the EU have the competence to do so?

This paper comprises five sections, with an introduction and a short overview of recent EU normative activities in the field of AI regulation. This is followed by a discussion of the intersections of AI and criminal law. For a clear and systematic view, this discussion is divided into the areas in which criminal law uses various forms of AI to assist in criminal prosecution (known as 'areas of

[9] See, eg, Daniela Rus 'The Robots Are Coming' (2015) 94 Foreign Affairs 2.

[10] Examples of such countries are Germany and Switzerland. For more on the situation in Switzerland, see, eg, Melinda Florina Müller, 'Roboter und Recht. Eine Einführung' (2014) 5 Aktuelle juristische Praxis 595. For Germany, see, eg, Sabine Gleβ and Thomas Weigend, 'Intelligente Agenten und das Strafrecht' (2014) 126 Zeitschrift für die gesamte Strafrechtswissenschaft 561.

cooperation') and areas where AI can appear as the 'perpetrator' of criminal offences (known as 'areas of collision'). The fourth section presents the views of the authors on what EU law should regulate in the future in this field, and in what manner, and the final, fifth section contains concluding remarks.

2 Recent legislative activities of the EU in the field of AI

The EU has been quite active with regards to civil law rules on robotics. In 2016 the European Parliament's Committee on Legal Affairs submitted a Report on the civil law rules on robotics to the EU Commission. Critics welcomed the adoption of such a document and commented that this was a timely effort by the Parliament, not only to remain in tune with the times, but also to take one step further and anticipate the future.[11] Among other things, this Report warned that the existing civil law framework was not sufficient to cover all the potential damage that could occur through the use of AI.[12]

Of course, fragmental regulation can be found in diverse documents. One good example of this is Directive 2016/680.[13] Article 11 of the Directive stipulates a prohibition of the decision based solely on automated processing, which is a provision that is analogous to the provision of Regulation (EU) 2016/679 (General Data Protection Regulation – GDPR).

The most recent regulatory act which should be mentioned is the Ethics Guidelines for Trustworthy Artificial Intelligence (AI) (hereinafter: Guidelines). This document was developed by a High-Level Expert Group on Artificial Intelligence (hereinafter: AI HLEG), which was established by the European Commission in June 2018. The AI HLEG consisted of 52 European experts who published the first draft of this document in December of 2018. After deliberations on over 500 received comments and suggestions,[14] the final version was published in April 2019. It should be noted that this was not the only initiative of this type at the international level, and that there have been over 84 attempts by various international associations to place AI into the frame of

[11] See, eg, Kirchberger (n 7) 195.
[12] European Parliament, Committee on Legal Affairs, Draft Report with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2013 (INL). For more on this, see, eg, Tjaša Zapušek, 'Artificial Intelligence in Medicine and Confidentiality of Data' (2017) 11(1) Asia Pacific Journal of Health Law & Ethics 109-113.
[13] Directive 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.
[14] Luciano Floridi, 'Establishing the Rules for Building Trustworthy AI' (2019) 1 Nature Machine Intelligence 261.

ethical guidelines.[15] The Guidelines are part of a broader European approach to artificial intelligence, which is reflected in the financial sector as well, through a 70% increase in annual investments into the innovative research programme Horizon 2020. Therefore, the EU will invest a total of EUR 1.5 billion between 2018 and 2020 into the development of artificial intelligence. This is also an attempt to reduce the gap in private investments in this field between the EU on the one hand, and the US and South Korea which are currently in the lead, on the other.[16] Other EU documents which were adopted include the Declaration of Cooperation on AI[17] (April 2018) and the Coordinated Plan on AI[18] (December 2018).

The purpose of the Guidelines is to promote reliable AI, based on legality, ethics and social and technical resilience, with an emphasis on a human-centric approach. The Guidelines determine the framework for the attainment of reliable AI, with the caveat that the legality of artificial intelligence is not considered.[19] The Guidelines provoked strong criticism soon after publication, especially because they failed to address potential high-risk areas, such as AI weapons. However, the principles listed in the Guidelines could be regarded as 'guidelines' for regulation which is recognised by the Panel for the Future of Science and Technology in the document 'The Ethics of Artificial Intelligence: Issues and Initiatives'[20] (for details, see 3.2 below).

This brief overview of the recent normative action of the EU in the field of AI leads to the conclusion that the field of criminal law (in its broader sense, ie substance, procedure and enforcement) has been completely disregarded thus far. The following sections will describe the fields considered to be significant and, in this regard, adequate for future regulation through EU rules.

---

[15] Brent Mittelstadt 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 Nature Machine Intelligence 1, available at SSRN: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391293> accessed 20 March 2020.

[16] See Coordinated Plan on Artificial Intelligence 'Made in Europe', available at <https://ec.europa.eu/digital-single-market/en/artificial-intelligence> accessed 20 March 2020.

[17] Available at <https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/eu-declaration-cooperation-artificial-intelligence> accessed 25 March 2020.

[18] Coordinated Plan on Artificial Intelligence, COM (2018) 795 final.

[19] High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy Artificial Intelligence (AI)' 7-9, paras 21-30, available at <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419> accessed 26 March 2020.

[20] European Parliament, Panel for the Future of Science and Technology, The Ethics of Artificial Intelligence: Issues and Initiatives, PE 634.452 – March 2020.

3 Areas where criminal law and AI intersect

In order to discuss the need for criminal law regulation of AI at the EU level, the scope of the relevant fields must first be determined. This refers to the fields in which criminal law and AI intersect. Only based on such an analysis can one determine whether the level of interaction of criminal law and AI is significant enough to merit more detailed criminal law regulation which would differ from the existing rules of criminal law.

A review of the literature shows that these two categories intersect at various points, depending on whether AI is used in a manner which is helpful to criminal prosecution authorities, or if it is used in a manner which can cause harm or other consequences to people and property. Here, a systematic approach has been chosen for clarity, and the presentation of these fields is divided into two groups: areas of cooperation (which implies the use of AI for the enhancement of efficiency in criminal prosecution), and areas of conflict (ie an overview of the criminal offences in which a form of AI can appear as the 'perpetrator'). This includes not only AI systems which are already in everyday use, but also those currently at the experimental stage, and which can realistically be expected to be in regular use in the near future due to their strategic significance. This analysis will also help determine the potential criminal law issues and future gaps related to the more active inclusion of AI in daily life.

3.1 Areas of cooperation

Scientists are currently developing and perfecting various forms of AI which will most likely play a role as tools to assist law enforcement authorities in the performance of their duties. For example, South Korea, which for a long time now has been trying to assert itself as the leading global force in technologies based on autonomous intelligence, has announced the introduction of robot prison guards in the near future. These robots should increase the level of security in prisons and relieve the burden of prison guards.[21] It can be expected that, if successful, such technology could soon enter European prison systems as well.

Relevant literature notes the problem by which the breakthrough of new scientific technologies, under the influence of the development of neuroscience and AI systems, could undermine the essence of the right to a fair trial. For example, Caianiello warns that the core of modern criminal proceedings lies precisely in the fact that each party to the proceedings has the right and the ability to challenge and contest the defendant's claims. However, as the

[21] For details, see 'Robot Wardens Are about to Join the Ranks of South Korea's Prison Service' (*BBC News*, 25 November 2011) <www.bbc.com/news/technology-15893772> accessed 27 March 2020.

mentioned author rightly points out, this right in recent years has been quite significantly limited by the new practice of the European Court of Human Rights since it has allowed its derogation in cases of overriding interests (eg preserving national security, combating terrorism, etc) and this right is essentially no longer absolute. Referring to the recent case law of Italian courts, Caianiello fears that a cultural unwillingness to adopt new technologies, in the absence of an adequate legal framework, will ultimately bring into question the protection of the defendant's fundamental rights and the 'dialectic approach to a criminal trial'.[22]

In discussions on the investigation of criminal offences and proceedings against perpetrators, one should bear in mind that the nature of criminal prosecution requires the immediate assessment of certain circumstances which are relevant for the further course of the proceeding. The accuracy of the assessment of such circumstances often determines the final outcome of the entire criminal proceedings. This applies especially to the preliminary stage of the proceedings (even before the formal initiation) when determining the existence of legal standards – such as reasonable suspicion – which are necessary to justify the continuation of the proceeding, and which can lead to the discovery and taking of key evidence (such as searches of persons and premises). However, if such actions are performed without sufficient grounds for suspicion or if this suspicion is not elaborated in an adequate manner, then this fact could make them illegal, thus leading to the exclusion of the collected evidence. This eventually affects the court's decision, which could result in exoneration by the court.

In order to avoid such a scenario, certain scientists are working on the design of software which could assist police to generate and logically connect the data relevant to the decision on whether or not to perform certain acts towards suspected criminal offenders. Such software entails complex algorithms which analyse the relevant data and suggest the most logical and most likely hypothesis, which can then be used to determine the reasonable grounds for suspicion for the performance of certain acts.[23]

---

[22] Michele Caianiello, 'Criminal Process Faced with the Challenges of Scientific and Technological Development' (2019) 27(4) European Journal of Crime, Criminal Law and Criminal Justice 267, 269.

[23] This idea is not only present in the context of criminal prosecution, but also in other areas. One example is the Collaborative Intelligence Spaces (CISpaces) software, which is being used as a tool to assist in the operation of US intelligence agencies. Available at <www.cispaces.org> accessed 26 March 2020. Another example is the Prototype Robotic Guard which is used by the US Border Control for the surveillance of borders. India has also announced the use of such an autonomous system. For more details, see 'AI Robots to Patrol India Border s Soon, Prototype to Come in December' (*Business Today*, 2 May 2019) <www.businesstoday.in/technology/news/ai-

A good example of such AI software is the program that is being developed by a group of scientists from Duquesne University in Pittsburgh, USA. Inspired by the fact that a large number of car searches resulting in the discovery of drugs are declared unlawful in court (due to the absence or insufficient grounds for suspicion), the scientists are developing a special system dedicated to assisting the police in the assessment of suspicion.[24] This system (which has not yet been named) functions in the following manner: after stopping a vehicle, the police officer verifies the data on the vehicle and the driver and enters certain parameters which serve as indicators of suspicion that the driver is transporting prohibited substances. These indicators are based on an analysis of tens of thousands of prior court records in drug trafficking cases and point to the criteria which the competent courts in these cases considered as being relevant for determining the legality of the search. Therefore, the police officer enters certain parameters into the system, such as: is there a strong smell of perfume emanating from the car (which can indicate that the driver wants to throw police dogs off the scent), what is the behaviour of the driver (is the driver nervous, is he or she sweating), is the car messy (which indicates a longer journey), were items removed from the trunk of the car which would otherwise belong there (eg the spare tyre – which indicates that the space was cleared for something else), is the official description of the route different from the actual route (eg the vehicle was rented and registered for a trip to destination A, but was travelling to destination B), has the driver previously been convicted for a criminal offence or misdemeanour, etc. The police officer enters such data into a laptop or even a smartphone through a standardised form, which allows for its quick completion. Based on the received data, the system provides the degree of probability that the car is transporting drugs, which gives the police officer a basis for further action.[25] Such a system carries significant weight in the US due to there being an issue of racial bias in police conduct in some states, and this could help prevent racial discrimination. While this is only an experimental project at this stage, it is realistic to say that similar systems will become commonplace in the near future due to the advantages they bring. At that time, an adequate legal framework will need to be developed as well, starting from the rules on the implementation of such systems and their maintenance and use, to rules on the admissibility of evidence collected by such means. This could be particularly interesting in Croatia, considering the possibility that the country

robots-to-patrol-india-borders-prototype-to-come-in december/story/342591.html> accessed 27 March 2020.

[24] Arthur Crivella, Wesley M Oliver, and Morgan Gray, 'Coding Suspicion' (JURIX 2018), available at <http://ebooks.iospress.nl/volumearticle/50850> accessed 27 March 2020.

[25] Arthur Crivella, Wesley M Oliver, and Morgan Gray, 'Reducing Subjectivity and Bias in an Officer's Analysis of Suspicion in Drug Interdiction Stops', Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, Montreal, QC, Canada, 17-21 June 2019, 254-255.

will soon find itself as part of the Schengen Area and that it will be extremely important to prevent the entry of various types of prohibited goods into the EU (as well as human trafficking). Therefore, this field merits the attention of EU law, at least in order to determine the basic principles of operation.

It should be noted that US criminal courts have already used algorithms in practice to assist the courts in assessing the risk of recidivism of defendants, which serves as one of the criteria for sentencing, ie for the replacement of prison sentences with alternatives, such as parole or community service. The sentencing process is generally considered one of the most controversial fields of criminal law, due to the existence of substantial differences in the sentencing policies of various courts. Although this issue exists in many systems worldwide, it carries special weight in the US with regards to the issue of racial discrimination.[26] Therefore, some states have introduced algorithms, such as the COMPAS system (Correctional Offender Management Profiling for Alternative Sanctions) in Wisconsin, which is designed to provide courts with a mathematical assessment of the risk of recidivism and eligibility for parole of the defendant. The complexity of the legal issues surrounding such an assessment of risk of recidivism is reflected in the *State v Loomis* case, which made its way to the Wisconsin Supreme Court. The plaintiff in this case (Mr Eric L Loomis) contested the consistency of the COMPAS system with the right to a fair trial. Loomis was denied parole on the basis of a COMPAS assessment, and he was sentenced to six years in prison, and an additional five years of special supervision after the completion of his prison sentence. In his appeal to the Supreme Court of Wisconsin, Mr Loomis claimed that COMPAS violated his due process rights because it used parameters which are relevant only to certain groups of defendants and because the methodology on which the system operates is kept confidential and cannot be subjected to a critical assessment by the defence.[27] The court ultimately rejected Loomis' arguments and upheld the verdict, but this case shows how the use of AI technology in the criminal judiciary can raise significant controversial questions, some of which (such as the right to a fair trial) are very significant in the context of European law as well. Considering that many European countries face the issue of sanctioning disparity, it is possible that some European systems will, under the influence of the expansion of AI trends, decide to implement programs such as COMPAS in the future, which will also require adequate legal regulation.

In the context of this discussion, we should also mention the use of AI systems during the interrogation of suspects. It is not always easy to determine whether

[26] For more details on this, see Darrell Stefensmeier, Jeffrey T Ulmer, and John H Kramer, 'The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male' (1998) 36(4) Criminology 763.
[27] *State v Eric L Loomis*, Supreme Court of Wisconsin, 2015AP157-CR, 13 July 2016.

a person is telling the truth during interrogation, even for experienced examiners. At the same time, this is a function of vital importance for the proceedings, given that the depositions of witnesses and the suspect can direct or even fully determine the further course of the proceedings. This applies particularly to situations where there are otherwise insufficient leads and evidence. There are certain assisting mechanisms that have already been used for questioning for some time now. The most well-known among these is of course the polygraph test, which is based on the recognition of certain psychological reactions by the respondent to the posed questions. In more recent times, the polygraph test has been joined by the somewhat more precise Guilty Knowledge Test, which consists of the application of technologies such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) and provides for the monitoring of blood flow in the brain. The scans obtained through these methods allow the recognition of certain physiological (not just psychological) reactions of the brain to the posed questions. The test is designed to determine whether or not the respondent recognises certain information connected to the criminal offence. If the information is recognised, then the brain will react differently than if the information is not recognised, and this will be visible on the screen.[28] This method is already in use in different parts of the world, and courts have, from time to time, affirmed that a positive result in such a test can be used as evidence of guilt in criminal proceedings.[29] A similar example exists in the EU, namely in Italy. The Italian court accepted two neuroscientific experimental procedures which tested the credibility of the victim and defendant in a situation where there was no other evidence. The court applied the Implicit Association Test (IAT) and the Timed Antagonistic Response Alethiometer Test (TARA), on the basis of which, with the help of an expert witness, it concluded that the defendant was guilty of sexual harassment.[30]

Although the described methodology can be very effective, it creates significant legal implications. For example, there are warnings that such interrogation methods violate a defendant's right to defend themselves by remaining silent.[31] It has been pointed out that such neuro-tests will fully replace the role of cross-examination and thus affect the nature of criminal proceedings in the spirit of what is called the Socratic approach to testimonial evidence.[32] Some authors take this one step further in their considerations and ask whether the future development of AI technologies for the interrogation of

[28] For more details on this, see Ciara Staunton and Sean Hammond, 'An Investigation of the Guilty Knowledge Test Polygraph Examination' (2011) 1(1) Journal of Criminal Psychology 1.
[29] See eg *Harrington v State,* Supreme Court of Iowa, 122/01-0653, 26 February 2003.
[30] Caianiello (n 22) 281.
[31] Kristen Thomasen, 'Examining the Constitutionality of Robot-enhanced Interrogation' in Ryan Calo, A Michael Froomkin, and Ian Kerr (eds), *Robot Law* (Edward Elgar Publishing 2016) 317.
[32] Caianiello (n 22) 282.

defendants will place the legality of the interrogation into question if the robot decides to use illegal techniques, such as torture. Such authors view the development of AI in this field as a potential severe threat to basic human rights, and they rightly raise the question of who will be liable in such cases, since modern criminal law has not taken a clear position on the liability for criminal offences committed by AI (for more on this, see the following sections).[33]

Since the development and implementation of AI technologies in criminal proceedings will certainly affect the EU region, we find that this area is significant enough to merit legal regulation at the EU level, and the establishment of basic principles and legal (as well as ethical) standards in this regard.

3.2 Areas of collision

While the previous section reviewed the most important areas in which AI can be used for the discovery of criminal offences, this section turns to the other side of the coin: situations in which AI can appear as the 'perpetrator' of a criminal offence. The term 'perpetrator' is used with reservation, because the criminal liability of AI has not yet been adopted (nor is it conceivable) in modern criminal law.[34] It should be noted that, due to the intensifying penetration of AI into various sectors, it is theoretically conceivable that the issue of AI as the perpetrator of a criminal offence will be viewed in relation to all (or at least most) criminal offences. However, we will not delve into such a deeper analysis at this point, but will rather focus on the areas where technologies with high levels of autonomy have already been introduced.

Among such sectors, the transport industry holds one of the most significant positions. Over the past several years, this sector has seen the intensified development of systems for steering assistance which reduce the need for drivers to actively participate in the driving process and transforms them into more of a supervisor, there just to correct potential errors or divergence from the desired driving mode by the automobile. Thus, serial automobile production has for years now also included assistance tools such as tempo mats with cruise control functions, the function of keeping the steering wheel steady in one lane, etc. It can be asserted that such equipment is no longer just reserved for the (most) luxurious automobile models, but can be found in automobiles meant for the

---

[33] See, eg, Amanda McAllister, who claims that in cases of torture during interrogations conducted by AI, there would be a separation of the *mens rea* (which can only be attributed to humans) and the *actus rea* (which would be attributed to AI in such cases). Amanda McAllister, 'Stranger than Science Fiction: The Rise of A.I. Interrogation in the Dawn of Autonomous Robots and the Need for an Additional Protocol to the U.N. Convention against Torture' (2017) 101 Minnesota Law Review 2554.

[34] The discussion on the issue of criminal law liability of AI goes beyond the scope of this paper. For more on this, see Bartosz Brożek and Marek Jakubiec, 'On the Legal Responsibility of Autonomous Machines' (2017) 25(3) Artificial Intelligence and Law 293.

general population.[35] While such automobiles are not yet in open supply in the EU, there are already pilot-projects in place. In some US states, however, it is notable that such (fully autonomous) vehicles participate in traffic, and there have also been cases where such vehicles have caused traffic accidents with consequences for other participants. This is illustrated in the next two examples. The first example relates to a situation in Arizona, where an 'Uber' vehicle, which was operating in 'full-autonomy mode', failed to detect a pedestrian who was crossing the street in an unmarked area. The vehicle continued to drive at full speed and hit the pedestrian, which resulted in her death. The second example relates to a similar accident (a chain reaction crash) caused by a 'Tesla' vehicle, which was also driving in autonomous mode. This accident also resulted in the death of one of the drivers in the other automobiles.[36] These cases have shown that autonomous vehicles, along with their undeniable advantages,[37] can also be dangerous. This raises the question of who bears the criminal law liability for such accidents? Is it the driver (who is actually not the driver), the owner of the vehicle, the seller, the programmer, or the manufacturer? Is there shared liability among these persons? What if these subjects are legal entities in a legal system which does not recognise the criminal liability of legal entities? Finally, what can these persons do, and is there anything they can do to prevent their criminal liability? Different legal systems provide different answers to these questions, and in many systems this issue has not been systematically discussed.[38] These are complex issues which require a deeper analysis and assessment of the basic concepts of criminal law, such as the principle of guilt, the limitations of liability

---

[35] This is what is known as full autonomy mode (autonomy level 4), which represents the highest degree of autonomy on a scale from 0 to 4 and is currently only in the experimental phase. It is predicted that this mode of operation will be a long-term replacement for human drivers and will thereby reduce the number of traffic accidents. For more, see National Highway Traffic Safety Administration, US Department of Transport, Preliminary Statement of Policy Concerning Automated Vehicles, 2013, available at <www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf> accessed 30 March 2020.

[36] Sam Levin and Julia Carrie Wong, 'Self-driving Uber Kills Arizona Woman in First Fatal Crash Involving Pedestrian' (*The Guardian*, London, 19 March 2018) <www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe> accessed 1 April 2020.

[37] The most commonly cited advantages are the removal of negative human factors, such as driver fatigue, slow reflexes, drunk driving, etc. They are also more precise in their reactions and allow for automobiles to be used by a population which would otherwise be unable to do so, such as the elderly, blind, deaf and other persons. For more details, see Frank Douma and Sarah Aue Palodichuk, 'Criminal Liability Issues Created by Autonomous Vehicles' (2012) 52 Santa Clara Law Review 1157, 1162.

[38] For more on this, see, eg, 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. For an analysis of Croatian law in this context, see Marin Mrčela and Igor Vuletić, 'Criminal Law Facing Challenges of Robotics: Who is Liable for a Traffic Accident Caused by Autonomous Vehicle?' (2018) 68 Collected Papers of Zagreb Law Faculty 465.

for negligence, and criminal law causality. At the same time, it is important not to frame criminal liability in a manner which will serve the purpose of prevention and social protection, but which will decelerate and hinder technological advances. Considering the need for the development and strengthening of the single EU market, with the significant diversity between the criminal law systems of individual Member States, it can rightly be stated that this area merits regulation through an adequate directive.

In addition, some authors point to the potential problems related to presenting evidence. They imagine a situation in which a partially autonomous car is called as witness after a traffic accident. Of course, such a 'witness' is in possession of a great amount of data which could have decisive influence in concluding about the existence (or absence) of the driver's guilt. At the same time, such a 'witness' creates several still unresolved dilemmas: how to access such information and how to ensure that it is trustworthy and admitted fairly?[39]

Based on the previous elaborations, the same logic can be applied to the medical field, or more precisely to the development and introduction of treatment methods performed by AI systems. The use of robotics with a certain level of autonomy is no longer a novelty in medicine. This is the case particularly in certain branches of medicine, such as diagnostics and surgery, where robot technologies are used to assist surgeons achieve a higher level of precision in their work, to prevent tremors, etc.[40] Several years ago, American scientists managed to develop a fully autonomous robot-surgeon system (called Smart Tissue Autonomous Robot, STAR), which performed independent surgery on a pig's abdomen. STAR performed the same procedure as a team of human surgeons who were operating on a different pig, and it performed better than the surgeons in terms of the selected method, timing and speed.[41] It is important to emphasise that the system itself selected the most appropriate method of operating, based on an assessment of the collected relevant data.[42] This raises the same questions which arose in the context of autonomous vehicles: who will be liable for the health consequences of a patient if the system makes an error in its selection or performance? How will that impact on the concept of criminal liability for medical errors, considering the significant differences which exist in

---

[39] Sabine Gless, 'AI in the Courtroom: A Comparative Analysis of the Machine Evidence in Criminal Trials' (2020) 51 Georgetown Journal of International Law 195.

[40] See, eg, David B Camarillo, Thomas M Krummel, and J Kenneth Salisbury, 'Robotic Technology in Surgery: Past, Present, and Future' (2004) 188 The American Journal of Surgery 2.

[41] See, eg, Eliza Strickland, 'Autonomous Robot Surgeon Bests Humans in World First' (*IEEE Spectrum*, 4 May 2016) <https://spectrum.ieee.org/the-human-os/robotics/medical-robots/autonomous-robot-surgeon-bests-human-surgeons-in-world-first> accessed 23 November 2019.

[42] See University of Maryland, 'Smart Tissue Autonomous Robot Raises the Bar on Surgery Precision' (*News Story*, 5 April 2018) <https://bioe.umd.edu/news/story/smart-tissue-autonomous-robot-raises-the-bar-on-surgery-precision> accessed 1 April 2020.

this field of criminal law in various states (some EU Member States provide for medical malpractice as a separate criminal offence and show a trend towards criminalisation, while other states treat medical errors as a component within the general offence of manslaughter and provide sanctions only for exceptional cases under the standard of gross negligence[43])? Considering the accelerated development of medicine and medical technologies, especially in Western Europe, it is necessary to establish certain minimal criminal law standards regarding causality and guilt in medicine concerning medical treatment with the use of autonomous systems.

The issue of causality and guilt is linked to some criminal offences which are considered typical intentional criminal offences if the perpetrators are humans. In this context, there are discussions regarding the development and use of autonomous weapons and the possibility that such weapons, due to an erroneous assessment of the target or the wrong selection of means, could cause (massive) consequences for non-military targets and thereby commit elements of war crimes. In this regard, one can rightly wonder whether the existing rules of command responsibility would be sufficient to convict superiors or whether there would be no responsibility at all due to the inability to prove the guilt of those superiors, regardless of the potentially large number of casualties. Recent military history shows that such a scenario is far from inconceivable. A very drastic and illustrative example can be found in the case of the downing of an Iranian civilian aircraft in 1988. The incident occurred when the American defence system for the protection of ships from airstrikes confused a civilian aircraft for a military plane and launched counter-aviation rockets, which led to the destruction of the aircraft and the death of 290 passengers and crew members, for which none of the responsible persons were ever convicted.[44] The development of such combat systems has continued to this day – for example, the US military has technologies such as the 15 Phalanx system which automatically and autonomously detects, assesses, follows and destroys different types of missiles which are used in attacks against ships. The Israeli military has systems from the 'Fire-and-Forget Weapons' system, such as the autonomous aircraft 'Israeli Harpy' which, once launched, finds, identifies and destroys radar transmitters on its own.[45] The development and perfecting of autonomous weapons is set as one of the strategic goals of many military superpowers in the world. Considering the possibility of far-reaching

[43] On this issue, see, eg, Miha Šepec, 'Medical Error – Should It Be a Criminal Offence?' (2018) 11 Medicine, Law & Society 47.

[44] Richard Halloran, 'The Downing of Flight 665: US Downs Iran Airliner Mistaken for F-14' (*The New York Times*, New York, 4 July 1988) available at <www.nytimes.com/1988/07/04/world/downing-flight-655-us-downs-iran-airliner-mistaken-for-f-14-290-reported-dead.html> accessed 2 April 2020.

[45] For details on this and similar technologies, see Heather M Roff, 'The Strategic Robot Problem: Lethal Autonomous Weapons in War' (2014) 13 Journal of Military Ethics 211.

consequences, this naturally raises the issue of criminal law liability and command responsibility in the context of war crimes. This issue is particularly significant from the perspective of European continental systems, which mostly reject the concept of liability for excess, such as joint criminal enterprise or conspiracy, and insist on proof of guilt of the superior.[46] Therefore, this issue should be regulated *de lege ferenda* at the EU level, which is currently a subject of discussion.[47] In this context, it is should be noted that the aforementioned Guidelines are being severely criticised by a number of scientists who claim that the Guidelines have failed to fulfil their main task as they do not take a position on autonomous weapons.[48]

Aside from the elaborated issues, criminology literature recognises other important areas where certain AI systems can appear as perpetrators of criminal offences with far-reaching consequences for people and property. Thus, there are warnings of the possibility of abuse which can occur due to the involvement of AI in the financial sector, especially through methods such as market manipulation,[49] price fixing, and collusion.[50] These methods imply the participation of an AI system which is designed to perform search tasks instead of people (Autonomous Trading Agent). Problems arise when such a system, with the ability to learn from its surroundings and received data, starts to emit information with the purpose of intentionally misleading the contracting party down the wrong path.[51] Some research has shown that such AI could master techniques of sending fictitious orders (which will never be performed) and concluding fictitious transactions, with the aim of defrauding good-faith third persons to gain profit. This could occur due to the fact that the AI is programmed to, among other things, find the most profitable business models. Therefore, it could probably happen that AI recognises the conclusion of fictitious transactions as the most profitable option and then operates accordingly. Furthermore, there can also be various types of illegal manipulations on the stock market, through the dissemination of false information on the value of

[46] For details on this, see Jack M Beard, 'Autonomous Weapons and Human Responsibilities' (2014) 45 Georgetown Journal of International Law 642.
[47] See also Roksandić Vidlička, Lelde Elīna and Ostapchuk (n 4) 281-282.
[48] Stuart Russell, 'Take a Stand on AI Weapons' (2015) 521 Nature – International Weekly Journal of Science 415.
[49] Market manipulation is any action and/or trade by market participants that attempts to influence market pricing artificially, with the intention to deceive and cause manipulative effects. See Michael P Wellman and Uday Rajan, 'Ethical Issues for Autonomous Trading Agents' (Noname manuscript) 4, available at <http://strategicreasoning.org/wp-content/uploads/2017/01/ethical-issues-autonomous.pdf> accessed 4 April 2020.
[50] Collusion is formal or explicit agreement among competitors with the purpose of earning greater than competitive profits. George A Hay and Daniel Kelley, 'An Empirical Survey of Price Fixing Conspiracies' (1979) 437 Journal of Reprints for Antitrust Law and Economics 439.
[51] ibid, 14.

shares by algorithmic trading agents.[52] All of the above raise the issue of adequate criminal law sanctions, considering that in most EU Member States economic crime is still predominantly linked to persons (human beings) under the traditional concept of guilt[53] (in such cases, primarily the intent of the perpetrator).

There are three potential models of criminal liability provided in literature. According to the first, AI should always be treated as a mere tool, which entails the liability of the programmer and the producer as kinds of indirect offenders. According to the second one, the emphasis is on the negligence of persons for AI, and they will be liable if it can be proven that the consequences were reasonably foreseeable. According to the third model, which is the least likely to occur in practice, criminal liability will arise pursuant to the direct liability model.[54]

Consequently, there should be considerations of a common supranational framework at the EU level, which would enable Member States to direct the future development of national criminal codes in this regard (either through redefining the concept of guilt and intent, or through a special section and the introduction of a separate criminal offence of abstract danger and shared liability of the manufacturer, programmer, seller and user of such AI systems).[55]

4 What issues should EU criminal law address?

After presenting areas considered relevant in the context of regulating *de lege ferenda* at the EU level, this section offers an overview of the question of which specific issues should be regulated by EU law and which direction should be taken in the development of national criminal law systems in this regard. The starting point is the assumption that the impact of EU law on national criminal law is less intensive than in other branches of law, primarily because of the nature of criminal law as a strictly national branch, which reflects the legal

[52] See, eg, Thomas C King, Nikita Aggarwal, Mariarosaria Taddeo, and Luciano Floridi 'Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions' (2020) 26 Science and Engineering Ethics, available at SSRN <https://ssrn.com/abstract=3183238 or http://dx.doi.org/10.2139/ssrn.3183238> accessed 4 April 2020.

[53] A systematic overview of the criminal law regulation of financial criminality in the EU Member States can be found in, eg, Judith van Erp, Wim Huisman, and Gudrun Vande Walle (eds), *The Routledge Handbook of White-Collar and Corporate Crime in Europe* (Routledge, Taylor & Francis Group 2015).

[54] Roksandić Vidlička, Lelde Elīna and Ostapchuk (n 4) 280.

[55] It can be noted that some national systems already provide for a considerably broad criminal framework, with regards to the criminal liability of the persons standing behind the AI, as well as the incrimination of special criminal offences of abstract endangerment. For more on this, see Mrčela and Vuletić (n 38) 483.

tradition specific to each national system.[56] Furthermore, the directives with criminal law content mostly regulate what is known as reversed vertical relations (in which the individual has duties towards the state and not the other way around), and therefore they exclude the possibility of direct action in the absence of timely implementation.[57] We also acknowledge that the previous EU regulation in the field of criminal law has been much more prominent in procedural than in substantive law, as well as the fact that regulations of substantive matters have been directed more towards special than general sections of criminal law.[58]

The basic precondition for criminal law regulation of the fields of AI operation with a higher risk of harmful consequences is the amendment of Article 83 of the Treaty on the Functioning of the European Union (TFEU). This Article regulates the areas of mutual interest which are of sufficient relevance to justify intervention. The list of nine areas of crime (Article 83 paragraph 1 TFEU) currently provided in this article is exhaustive[59] and there cannot be any influence outside this defined scope. The preceding sections, however, show that various risks can be expected in fields such as medicine and medical criminal law, transport or war crimes. Therefore, the list from Article 83 paragraph 1 should be expanded *de lege ferenda.* In this sense, it would be sufficient to add a formulation such as 'criminal offences connected to autonomous intelligence' because this would cover other potentially dangerous areas which may not be discernible at the moment.

It is evident from the previous elaborations that the existing questions concern substantive, procedural and enforcement law. In the substantive sense, they relate to the rules on minimal substantive standards of criminal offences which all countries should enact in order to protect society from the dangerous acts of AI. In the procedural sense, they relate to rules on the limits of admissibility in the use of specific AI systems in criminal proceedings. The rules relevant for enforcement law should regulate the contribution of various forms of AI in the process of the enforcement of prison sentences, house arrest and other prison alternatives, as well as monitoring compliance with security

[56] Some authors advocate the adoption of a 'European Criminal Code' and find that such a development is inevitable in the near future. See Helmut Satzger, *International and European Criminal Law* (CH Beck – Hart – Nomos 2012) 43.
[57] ibid. See also Case C-457/02 *Antonio Niselli* ECLI:EU:C:2004:707 para 29; Case 80/86 *Kolpinghuis Nijmegen* ECLI:EU:C:1987:431 para 13, etc.
[58] This is logical, considering that the general sections regulate the basic principles and institutes, and thereby reflect a specific national tradition in a greater measure than the special sections of criminal codes. However, there are authors who warn of the need for the harmonisation of the general section of criminal law at the supranational level. For more on this, see Andre Klip (ed), *Substantive Criminal Law of the European Union* (Maklu Publishers 2011).
[59] See also Satzger (n 56) 74.

measures like criminal law sanctions determined in a judgment at the end of court proceedings (such as restraining orders against persons or places), etc.

From the perspective of substantive law, it would be difficult (or perhaps even impossible) to oblige the Member States to amend the general section of criminal law by adopting concepts which would allow the determination of guilt outside the limits of the foreseeability of the consequence, such as the common law concept of joint criminal enterprise. This would be unacceptable in those countries of the continental tradition which are still firmly based on the principles of legality and guilt and reject any form of objective liability.[60] Therefore the draft of a future directive should be aimed exclusively at the modification of the special section of criminal law, by obliging the states to incriminate new forms of criminal offences of what would be perceived as abstract danger, for which criminal law protection in those occupations which include the use of AI systems would be extended and pulled back to earlier stages. At the same time, the general adoption of the concept of criminal liability of legal entities should be insisted on. This would create conditions for the adequate sanctioning of all situations where AI systems are placed on the general market without maximum possible measures for the control, protection and prevention of consequences. These would, of course, be blanket norms of criminal law, because they would refer to other provisions to regulate each individual industry branch in this sense. Special attention should be given to the issue of the development and use of autonomous weapons. In this regard, modification of the concept of command liability in the appropriate direction should be considered, taking into account the massive number of casualties which such weapons can cause.[61]

From the perspective of procedural and enforcement law, there should be clear, common standards for the use of AI in these sectors. Particular emphasis should be placed on the need to protect human rights and the due process rights of suspects, as well as the mandatory application of all measures for the prevention of torture and other forms of inhumane and demeaning conduct, especially in proceedings with arrested persons and detainees. It may become necessary to establish appropriate standards regarding the evidentiary value of evidence collected by means of AI systems, with special emphasis on the issue

---

[60] For details on this, see Christoph Barthe, *Joint Criminal Enterprise (JCE). Ein (originär) völkerstrafrechtliches Haftungsmodell mit Zukunft?* (Duncker & Humblot 2008).

[61] It should be noted here that this will by no means be an easy task, bearing in mind the disagreements which currently exist around certain types of command liability in the context of causality and guilt (for example, the types based on the 'Yamashita' model). For more on this, see Petar Novoselec, 'Materijalne odredbe Rimskog statuta i njihova implementacija u hrvatskom kaznenom zakonodavstvu' in Ivo Josipović, Davor Krapac and Petar Novoselec (eds), *The Permanent International Criminal Court* (Narodne novine 2001). See also William A Schabas *The International Criminal Court: A Commentary on the Rome Statute* (OUP 2010).

of determining reasonable suspicion arising out of data processed by AI and which can be used as grounds for the continuation of the investigation process.

5 Conclusion

Artificial intelligence is, without doubt, a new technological force that is in full swing, and which will transform society and life as we know it. There are many advantages of such a development, starting with the fact that AI creates great opportunities for economic progress and growth, as well as for enhancing the quality of life and quality of services, and also for reducing the harmful effects of industry, the emission of greenhouse gases, etc. Therefore, it is of high strategic importance for Europe to 'catch up' with the developed industries of the US and Asia. At the same time, it is imperative to find the right approach to balance the broad advantages with the potential risks. In this sense, establishing an adequate legal framework is of key importance.

As emphasised in the introduction, the purpose of this research was to answer two questions: are there fields in criminal law which are ripe for regulation and is the EU the appropriate level for this kind of regulation? In our opinion, the previous analyses have proven that the answer to both these questions is positive. The overview of comparative literature and case law (at least in some countries) shows how AI can be (and already is) in close relation to both substantive and procedural criminal law. In both fields, there is considerable need to establish new concepts, since traditional ones will soon not satisfy daily requirements. We believe that EU law is the proper tool for supranational intervention, especially given that AI law (and also AI criminal law) is still in its infancy and therefore very suitable for harmonisation.

The normative activity of the EU in this area is of great significance, because it can contribute to the harmonisation of national systems and to the harmonised development of the single market. Naturally, while we are aware that such normative activity is primarily focused on the branches of private law, we nonetheless find that it should not circumvent public law, and criminal law in particular.

Based on the previous elaborations, it can rightly be concluded that criminal law is facing its third revolution in the course of two decades, where 'revolution' here means the modification of fundamental principles, reasoning and approach to the concept of criminal offences and their perpetrators. The first such revolution occurred with the development and global adoption of the concept of the criminal liability of legal entities. The second revolution took place when criminal law developed completely new concepts to efficiently respond to criminal offences which occur in the virtual world, which lead to the global acceptance and incrimination of cybernetic criminality. Both revolutions were conditioned by the newly developed social needs of the given moment.

Criminal law is as dynamic as the development and progress of society. It follows this progress and is shaped by social needs, ie the need for social protection. In this sense, it is important to take a clear position regarding the development of potentially catastrophic autonomous technologies, such as autonomous weapons. Thus, in the inevitable next stage of the development of criminal law (the third revolution), it will be necessary to design and implement concepts that cover criminal liability for the acts of AI. It is important for the EU to recognise this in a timely manner.