COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 4 September 2020 CDPC(2020)3Rev

EUROPEAN COMMITTEE ON CRIME PROBLEMS

(CDPC)

FEASIBILITY STUDY ON A FUTURE COUNCIL OF EUROPE INSTRUMENT ON ARTIFICIAL INTELLIGENCE AND CRIMINAL LAW

Document prepared by the Working Group
on AI and Criminal Law
with Dr. Sabine Gless as General Rapporteur
and in co-operation with the CDPC Secretariat

www.coe.int/cdpc | DGI-CDPC@coe.int

TABLE OF CONTENTS

Introduction

The fact that robots have become part of our daily lives raises many and different practical and legal novel issues including in the sphere of liability, especially those in criminal law. This is especially noticeable with driving automation, whether employed in cars, aircraft or drones, but it also applies to robots operating, for instance, in the field of medicine or financial transactions, or helping to care for the elderly.

Facing the prospects of the digital age, state authorities and international organisations are working hand in hand as there is yet no common legal framework or international cooperation governing this situation. To be able to bring the promised benefits to the people, some countries, however, have already adopted specific regulations on driving automation that deal with the use of artificial intelligence (AI) for specified use and thus addresses liability matters for the use of AI.

Different national standards and legal regulations, however, could endanger a common leap forward in Europe as well as an adequate protection of individual interests and the maintenance of legal certainty.

The Council of Europe is unusually well placed to provide assistance in common standard setting with its successful track record in defining benchmarks and offering harmonised approaches for human rights protections, cybercrime issues and mutual legal assistance, on both a collective and individual level, based on a stable network of member states' criminal justice cooperation, committed to effective enforcement measures while complying with fundamental human rights, in particular defence rights of persons prosecuted for alleged crimes and the rights of victims.

The European Committee on Crime Problems (CDPC) has been entrusted by the Committee of Ministers with responsibility for overseeing and co-ordinating the Council of Europe's activities in the field of crime prevention and crime control[1]. It has made the substantive criminal law challenges posed by advances in robotics, AI and smart autonomous machinery, capable of causing physical and material harm or even death, independent of human operators, a priority[2]. On 28 November 2018, the CDPC held a thematic session in Strasbourg on AI and criminal liability, for the purpose of (i) examining and ascertaining the scope and substance of relevant national criminal laws and international law pertaining to the use of automated vehicles (or other AI deployment) (ii) determining in what circumstances certain types of conduct are or should be covered by criminal law in relation to the delegation, division or assignment of tasks, functions and behaviours to automated technologies, and the possible cross-border relevance, (iii) examining the scope and substance of an international legal instrument to provide common standards for the criminal law aspects of automated technologies, in particular with regard to driving automation.

At the thematic session, the CDPC recognised the importance of this topic and decided to set up a restricted working group of 15 representatives of member states supported by a number of scientific experts (hereinafter referred to as the "Working group")[3]. The Working group was tasked with (i) taking stock of existing regulations, (ii) identifying future challenges related to the development of AI to be addressed in the criminal law field, with a particular focus on criminal liability and license conditions for the marketing and use of items equipped with AI and (iii) making proposals for possible action and standard-setting activities in this field.

[1] The CDPC identifies priorities for intergovernmental legal co-operation, makes proposals to the Committee of Ministers on activities in the fields of criminal law and procedure, criminology and penology, and implements these activities. The CDPC may prepare conventions, recommendations and reports.

[2] CDPC – List of decisions – 73rd plenary meeting, CDPC (2017) 27, para 3, p. 2.

[3] CDPC – List of decisions – 75th plenary meeting, CDPC (2018) 21, para 3, p. 2.

An initial meeting of the Working group was held in Paris on 27 March 2019, the first objective being to compile a questionnaire so that key information could be gathered from member states at national level, identifying possible existing gaps in criminal law and criminal law solutions already in place, and the second objective being to assess the necessity for an international instrument on AI and criminal law. The questionnaire was sent to member states in May 2019, based on the example of driving automation. An evaluation of the replies to the questionnaire was published on 21 October 2019, after replies were received from 36 member states (Andorra, Armenia, Austria, Azerbaijan, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Italy, Moldova, Monaco, Montenegro, North Macedonia, Norway, Latvia, Lithuania, Luxembourg, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and Ukraine).

Following on from these projects, at the 77th plenary meeting of the CDPC held in Strasbourg from 3 to 6 December 2019, the Working group was instructed to "*carry out a feasibility study identifying the scope and the main elements of a future Council of Europe instrument on AI and criminal law, preferably a convention"*[4].

This document therefore aims to answer the following questions: should an ad hoc Council of Europe committee of experts be set up to prepare a draft instrument setting common criminal law standards on different relevant issues raised by vehicles driving autonomously (or other AI deployment)? If so, what specific legal topics, identified as problematic at the international level, should this group address?

Following a brief reminder of the main issues at stake in the development of AI in relation to criminal liability (see Chapter 1.) and the regulations in this area (see Chapter 2.), the advisability of adopting a legal international instrument on AI and criminal liability (see Chapter 3.) and the legal issues to be addressed in such an instrument (see Chapter 4.) will be considered.

[4] CDPC – List of decisions - 77th plenary meeting, CDPC (2019) 23, para 7, p.4.

1. The phenomenon of AI and its criminal law impact

1.1. Data analysis, automation and self-learning capacity

In situations where, in the past, humans have taken decisions and acted accordingly, today technology is used to replace human performance with automated IT systems which operate with the help of algorithms. These systems can – based on machine-learning – collect and process large quantities of data in order to act autonomously, and thus can replicate and assume decision-making functions previously performed by humans. This new capability of processing a plethora of data, based on algorithms and machine learning, enabling the detection and analysis of patterns to generate autonomous activity and possibly new decision-making rules automatically, not offered by humans, often is referred to as Artificial Intelligence (AI), see also infra 4.1.3.

AI uses techniques derived mainly from statistics, computer science and cognitive psychology to replicate tasks traditionally left to humans. Autonomous learning systems evolve by adapting to the information transmitted by their sensors or updates, with designers determining and adjusting only the initial settings and the overall objective to be achieved in an optimal manner.

This technology has rapidly found its way into numerous, ever more complex areas such as fraud prevention, risk analysis including in the legal, financial and insurance sectors, consumer intelligence gathering, targeted advertising, medical equipment, and automated driving both on the ground and in the air. The proliferation of digital services and AI systems has brought significant benefits, offering people greater convenience and efficiency across a wide range of fields and activities. Yet, as it stands today, AI lacks a holistic understanding of situations and can only accomplishes limited tasks. Again, the prominent example for such 'narrow AI' is driving automation.

1.2. Intelligibility and margin of error of AI driven systems unknown to humans

The beneficial solutions promised by AI applications will not be immune against errors: A car running on driving automation may miss a traffic sign and run over a pedestrian. A tool monitoring financial transactions for fraud prevention might flag out a payment pattern that is legal and doing so place a human under suspicion of criminal wrongdoing when all the person did was paying her/his bills. A medical device driven by AI could mistake malignant cancer for begin tumour and miss to point it out to the doctor. AI solutions have their specific strengths, but also weaknesses which can develop into risks, or even cause harm and death. Nowadays humans cannot be expected to foresee all possible outcomes of an employment of AI.

Liability, and in particular criminal culpability, however requires foreseeability of the risk of harm. The problem of the (un-)predictability of the outcomes achieved employing AI driven systems has quickly come to be framed in terms of a possible responsibility gap, as, with the advent of deep learning technologies, opaque systems have developed that humans can neither understand nor explain from the initial models, leading to semi-automated or automated decision-making that is unintelligible.

This issue is all the more significant in that some important aspects of human decision-making processes cannot be automated or assumed by AI. Human and algorithmic decision-making are fundamentally different, with distinctive consequences and errors. The weaknesses, limitations and boundaries of algorithm-based decision-making mean there are inherent risks in this new technology, which have to do with the reliability of algorithms as tools, human perception and interpretation of their implementation and results, and the acceptance of a decision outcome.

The issue of allocation of responsibility for the risks created by the development of AI systems is central, therefore, as the decision about what risks we ought to take (e.g. on public streets or in fraud prevention) and how we allocate responsibility is eventually a political one, involving the whole of society.

1.3. AI, criminal law and human rights

With the implementation of AI systems and the occurrence of the first incidents and accidents, it became apparent that, in the area of criminal justice, not only the allocation of responsibility, but also respect for certain fundamental rights, especially judicial rights, was a particular issue, as technological development had the potential to jeopardise the effective enjoyment of these rights in the near term or even, in some countries, now.

In particular, respect for the right to a fair trial, as enshrined in Articles 5, 6 and 7 of the European Convention on Human Rights (ECHR), together with respect for private life, safeguarded by Article 8 ECHR, are brought into play when data collected and used by AI systems may be offered as evidence in criminal prosecutions.

In the member states of the Council of Europe, criminal law is generally considered to relate to the conduct and intentions of humans only, whether natural persons or individuals acting on behalf of entities (corporate liability). Because they are so complex, hi-tech systems are apt to be misunderstood or insufficiently understood by designers, manufacturers, regulators and users, obliging all the parties concerned to be aware of their respective rights and duties. In this context, criminal liability deriving from situations where AI systems cause serious harm to humans needs to be made clear by unambiguous procedural safeguards and well-defined rule of law principles.

1.4. Criminal liability and AI: the example of driving automation

The example of driving automation is particularly telling where criminal liability for AI is concerned; used by the Working group in its discussions, it will be briefly repeated here to illustrate the main issues in this area.

Of the many road accidents recorded, some involve cars in autonomous mode[5] and of these, some have even been fatal. When an operator is present in the passenger compartment of a car running on driving automation at the time of the incident, the question arises of who is criminally liable. While traditionally, excessive speed and the consumption of alcohol or drugs by the driver were often relevant factors, they are no longer the only pertinent ones in such matters. Even if the technology used for driving automation appears defective, there is no simple answer to the question of criminal liability when AI systems are employed. The issue is crucial, however, because a clear allocation of liability and its enforcement by courts ensures reliable and peaceful co-operation within society.

However, liability strictly speaking is not the only concern. Highly automated cars generate valuable data when driving, that can be useful in many respects, including for law enforcement and criminal investigations. If an automated drive ends in an accident, the question, for instance, arises of whether these data can be used as evidence in a criminal proceedings, in particular against the human driver, and how to test the credibility of the systems that generate the data or the reliability of such data as evidence. Traditional evidence rules, however, may not be designed to meaningfully test the reliability and credibility of this new digital evidence.

[5] The industry distinguishes between different levels of driving automation (SAE standard J3016_201401). At level 2, a car can execute dynamic driving tasks, but the driver must monitor and overrule the system, if necessary. At level 3, a driver no longer needs to monitor the system when activated but must respond to a takeover request. Level 4 automation is used for various 'mixes' of highly automated and fully automated driving depending on the focus; it especially covers situations in which the driver does not respond to a take over-request (TOR) and the car is expected 'minimize' the risk resulting from this situation. Level 5 envisages autonomous driving without a human driver. The manifold forms of driving assistants draw on miscellaneous forms of machine learning-concepts for varied functions (e.g. lane and distance keeping, parking, infotainment, drowsiness detection). In order to be able to function on the road, driving assistants must be adaptive, i.e. capable for a specific operation based on autonomous data obtainment and evaluation. Thus, driving assistants embody AI risks (autonomy, connectivity and human-robot-interface) on different scales.

2. Existing criminal law on AI

2.1. Current legislation in Council of Europe member states

According to the answers to the questionnaire sent out by the Working Group and completed by the member states, only a few have prepared or already adopted general legislation which may affect criminal liability when humans hand over the steering wheel to driving assistants, in particular with regard to requirements for negligence. A larger number of states have adopted regulations for driving automation pilot projects, focusing legislative efforts on the implementation of general technical standards for special permits allowing automated driving, and on regulating the functions that such highly automated vehicles must offer. Member states which have chosen to regulate test driving only often make decisions on a case by case basis. It is rather difficult on that basis to draw a general conclusion about a common direction for such regulation[6].

France is an example of a country that has adopted provisions on the use of automated driving. The entry into service of vehicles with delegation of driving has been authorized in France since the entry into force of ordinance No. 2016-1057 of 3 August 2016 relating to the testing of vehicles with delegation of driving on public roads, subordinating any test to the issuance of a public authority authorisation. Law n ° 2019-486 of 22 May 2019 relating to the growth and transformation of companies (PACTE) amended the aforementioned ordinance, in order to broaden the scope of the tests, while specifying its legal framework by providing in particular the applicable criminal liability regime. Accordingly, the driver of a vehicle is no longer criminally liable for offences committed while driving the vehicle if the delegated driving system, which the driver has activated in accordance with the operating conditions, is in operating mode and informs the driver in real time that it is in a position to observe traffic conditions and to instantly perform any manoeuvre in his or her place. Criminal liability is transferred back to the driver, however, if a take-over request is issued and after a specified time for regaining control of the vehicle. Criminal liability is likewise transferred to the driver if he or she ignores the obvious fact that the conditions for using the driving delegation system, as specified for the purposes of the test, have not been met or are no longer met. Also, if the driving delegation system has been activated and is functioning according to the operating conditions, the holder of the authorisation for testing a vehicle driving autonomously is liable for the payment of any fines resulting from violations committed while driving, and criminally liable for any offences involving unintentional injury to life or limb if the driving caused an accident resulting in bodily injury, where a fault in the implementation of the driving delegation system is established.

Another interesting example is Germany which has adopted an amendment to its traffic law in 2017, according to which drivers are released from the obligation to monitor and pay attention when a licenced automated driving system is activated compliant to all rules[7]. However, the consequences of this regulation for a driver's held criminally liability using a car with Level 3 autonomy when an incident happens, yet, has not been spelled out clearly. Noteworthy, automated vehicles have to be fitted with a data recording device[8].

The case of Germany highlights several problems, among them for instance: How to allocate liability among the human driver and a driving assistance system and possibly prove that an accident was

[6] CDPC (2019) 17, p. 6.

[7] Cf. § 1b of the German Road Traffic Act: (1) The vehicle driver may turn his attention away from traffic and vehicle control while driving by means of highly or fully automated driving functions according to § 1a; in doing so, he must remain ready and alert so that he can fulfil his obligations under paragraph 2 at any time. (2) The vehicle driver is obliged to take control of the vehicle again immediately, 1. if the highly or fully automated system prompts him to do so, or 2. if he recognises, or due to obvious circumstances has to recognise, that the requirements for the intended use of the highly or fully automated driving functions are no longer fulfilled.

[8] Cf. § 63a of the German Road Traffic Act.

caused by a system failure and how to obtain this evidence from the vehicle manufacturer? Or the question of whether society can accept that, in some cases, it might be that no one can be held criminally liable, even if an automated vehicle causes a death?

2.2. International initiatives on AI and criminal law

At present, there are no regional or international regulations on AI and criminal liability. Many organisations and institutions are looking at the subject of AI in general, however, and addressing the question of civil and criminal liability in this area, while several universities and committees are studying the above issues in depth.

In particular CoE e.g. Council of Europe's (CoE) European Ethical Charter on the Use of AI in Judicial Systems of 2018,[9]

In the case of the European Union, the European Commission has set up a high level expert group, which published guidelines in April 2019 for a reliable AI, listing seven key requirements: (i) human agency and oversight, (ii) technical robustness and safety, (iii) privacy and data governance, (iv) transparency, (v) diversity, non-discrimination and fairness, (vi) societal and environmental wellbeing, and (vii) accountability.

On 19 February 2020, the European Commission adopted a white paper entitled "*On Artificial Intelligence - A European approach to excellence and trust*", supporting a regulatory and investment-oriented approach with the twin objective of promoting uptake of AI and addressing the risks associated with certain uses of this new technology. The purpose of this white paper is to set out policy options on how to achieve these objectives, with the Commission inviting *inter alia* member states and other European institutions to react to the options offered and to contribute to the Commission's future decision-making in this area.

In this white paper, the Commission underlines that it is vital that European AI is grounded in European values and fundamental rights such as human dignity, privacy protection and fair trial, and concludes that a common European approach to AI is necessary to reach sufficient scale and avoid fragmentation of the single market, recognising that essential work on AI is currently ongoing, including at the Council of Europe.

As regards the issue of liability for faulty products, which is the only element of liability studied by the European Commission, the Commission recommends adjusting or clarifying existing legislation in this area, or even introducing new legislation specifically on AI, with mandatory requirements in high-risk AI applications, in order to ensure effective judicial redress for parties negatively affected by AI systems and to ensure legal certainty and competitiveness for companies marketing their AI-based products in the European Union. The White Paper, however, identifies the same issues in this area as those mentioned above in relation to criminal liability, notably the need to define AI, the opacity of AI systems and the question of how obligations are to be distributed among the economic operators involved.

2.3. The CAHAI (Ad hoc Committee on Artificial Intelligence)

Pursuant to its terms of reference, the CAHAI is to "examine the feasibility and potential elements on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law" […] in co-ordination and consultation with other intergovernmental committees working on the subject.

[9] Accessible at <rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

Following its first meeting (November 2019), the CAHAI is carrying out a broad mapping exercise of all the work and instruments applicable to artificial intelligence, including European Convention on Human Rights and artificial intelligence impact (risks and opportunities) on human rights, rule of law and democracy. The mapping will allow identifying possible gaps in such instruments but also common, transversal principles for artificial intelligence design, development and application.

The CAHAI will consider possible options in terms of legal instruments and tools, however a clear stance on whether it should orient its work towards the preparation of a legal binding instrument can only be taken once the implications are fully analysed (November 2020 at the very earliest).

There are clear expectations from Council of Europe's member States of coordination of CAHAI work with other international organisations, in particular the European Union (EC, FRA, EDPS), UNESCO or OECD in order to promote synergies and avoid any duplication. In this respect, the Council of Europe and these organisations are participating in each-others' respective committees on artificial intelligence.

3. A legal international instrument on AI and criminal justice

3.1. Assessment of the need for a legal international instrument

Review of the replies to the questionnaire shows that, whether member states have adopted specific regulations on AI and criminal law or not, they typically remain rooted in traditional notions with regard to criminal liability regimes. The Working group has accordingly identified two tendencies[10]:

- either all responsibility remains with the human driver, which can lead to drivers being held liable in a way that member states themselves consider might be unfair, particularly in cases where the accident occurs when the vehicle is in automated driving mode and the driver has complied with all the rules[11];
- or drivers may divert their attention from the traffic situation and hand over the vehicle to driving assistants as long as they use the automated driving functions properly and are ready to respond to a take-over request at any time. In cases such as these, the Working Group has identified a legal vacuum, firstly where, despite compliance with all the rules, a fatal accident occurs and, secondly, as regards the degree of negligence required on the part of the driver to trigger his or her criminal liability, since holding drivers automatically liable if they fail to comply with the instructions given by the AI system through lack of time or lack of technical knowledge could also be considered unfair.

The Working Group accordingly noted that member states agree on the compelling necessity for new regulation, at the latest when "cars drive by themselves"[12]. Given automated decision making, the highly complex nature of machine learning processes and the fact that the human driver can no longer be held responsible for all driving activity, existing legislation does not sufficiently cover liability issues anymore.

Although these issues are crucial elements for member states' criminal justice systems, which is exclusively a matter of each individual country to decide, it appears essential that national regulations should be developed within an international and collaborative framework, for several reasons.

Firstly, as cars and humans cross borders the drafting of an international legal instrument on these issues will make it possible to provide member states with some common basic pointers so that consistent legislation can be developed across Europe. This need for consistency is all the greater as certain AI applications, and in particular driving automation, affect the ability of citizens to move around Europe, where state borders are crossed with ease and the technology does not necessarily register

[10] CDPC (2019) 17, p. 5.
[11] In particular, where the accident is due to a fault in the AI system, in terms of data collection or the rules applied.
[12] CDPC (2019) 17, p. 6.

the change in legal framework and sovereignty. It is not a question of devising a whole new system of liability that would overturn the criminal law of each member state, but rather of agreeing on a general framework for criminal law and AI deployment within which state-wide regulations could be developed.

Secondly, an international framework within national legislation on AI and criminal law will bring a degree of legal certainty to European citizens and users of the justice system as well as to the industry providing the technology, thus supporting its development in compliance with fundamental rights rather than hampering it with incompatible legislation. A common instrument assists free movement across national borders and prevents long-winded case-by-case judicial decisions. The view that it is too early for regulation in this area as the technology is still in its infancy, must be weighed against the benefits of working on an international framework now which can provide guidance when developing this technology. It is important to bear in mind that criminal liability issues are already cropping up in relation to the use of AI in automated vehicles and developing an international legal instrument in this area will take at least several months.

Thirdly, an international legal instrument of this kind will facilitate better co-operation between states on the different issues raised by AI employment in the area of criminal justice. Such co-operation would appear to be essential, particularly in matters relating to taking evidence from abroad (in regard of cases related both to individuals and legal persons) or transfer of criminal proceedings. The replies to the questionnaire show that all member states recognise the need for data monitoring and storage[13]. Most state authorities are already using various digital analysis tools to enforce safety on public roads (e.g. speed cameras and radar guns, digital breathalysers, automatic number plate recognition, smart tachographs for trucks, anti-alcohol engine locks, GPS positioning of vehicles). Without, however, addressing ongoing concerns about the exchange of such data when driving automation becomes a standard, more unresolved question will arise concerning for instance, privacy protection, respect for fundamental rights when using such data as evidence or reliability of such evidence.

3.2. The potential of the Council of Europe to pave the way for the adoption of an international legal instrument on AI and criminal law

The Council of Europe was set up to secure democracy based on the freedom of the individual and to prevent a recurrence of the mass human rights violations committed during the Second World War. The Council covers all major issues facing the European countries, other than military defence, and aims to promote democracy, human rights and the rule of law, and to develop common responses to political, social, cultural and legal challenges in its member states. The task of promoting adherence to the rule of law becomes particularly urgent when developing AI systems. By virtue of the work it has already done in this area, the Council of Europe is now a leading intergovernmental organisation on AI and criminal law.

The preparation and adoption of a legal instrument dedicated to AI in criminal justice will provide a means of dealing efficiently and consistently with alleged crime in this area and related problems which are slowly beginning to appear, including in Europe, by paving the way for the development of national legislation to fill any identified gaps according to common international standards. It will also help to ensure and promote greater international co-operation on these new topics, sending a strong signal to providers of AI driven systems about the need to develop this technology in a way that respects internationally protected fundamental rights. An international convention could be one of the tools with which to start negotiations to emphasis the urgency of the situation.

In view of the Organisation's broad membership (47 member states), a Council of Europe instrument would potentially have a powerful impact. In accordance with its human rights–based approach, a Council of Europe instrument on AI and criminal law would provide a link to the European Convention

[13] CDPC (2019) 17, p. 6.

on Human Rights, which places a positive obligation on each party to protect its citizens against human rights violations. In addition, the adoption of a Council of Europe instrument on AI and criminal law, having regard to existing international conventions (such as the 1949 Geneva Convention on road traffic or the 1968 Vienna Convention on road traffic[14]) and building on work under way in other international institutions, may prompt further action by the European Union and could provide inspiration for other international instruments.

4. Key elements of an international Council of Europe instrument on AI and criminal law

Using the traditional structure of Council of Europe instruments, the following issues, identified as important for Council of Europe member states and third countries to address, could provide a focal point for future negotiations with a view to adopting an international instrument on AI and criminal law.

4.1. Purpose, scope and definitions

4.1.1. Purpose of the instrument

As set out above and in the light of the analysis of member states' replies to the questionnaire on AI and criminal justice (using the example of driving automation), four objectives have been identified, namely:

1. To establish an international framework for the development of national legislation on criminal law issues in relation to AI (more particularly regarding criminal liability in the context of driving automation);
2. To encourage member states to take into account the legal issues in the area of criminal law and AI by addressing problems through legislation, using common normative principles;
3. To anticipate the evidentiary and other legal problems already identified in relation to criminal liability and AI and to ensure fair trial-principles as well as effective international co-operation in this area; and
4. To ensure the development of AI systems in accordance with the fundamental rights protected by Council of Europe instruments.

These general objectives may be discussed, modified and supplemented by member states.

4.1.2. Scope of the instrument

The aim here is to define the precise scope of a future Council of Europe instrument on AI and criminal law and not to encroach on the domain of other bodies working on other related issues.

Accordingly, and in the interests of clarity, issues relating to civil and administrative liability resulting from the use of AI will not be addressed as part of the process of developing a Council of Europe instrument on AI and criminal law, as these are separate issues and come under a different legal framework.

Similarly, the use of AI by member states' armed forces will not be covered by this future instrument, as the issue of citizens' criminal liability only arises in civilian applications of AI.

[14] The member states of the Vienna Convention voted for an amendment which adds two definitions (of an automated driving system and of dynamic vehicle control) as well as an article which, in essence, specifies that an obligation for any moving vehicle to have a driver is considered satisfied when the vehicle uses an automated driving system which complies with national and international technical regulations and national regulations govern its operation / movement on the road.

It appears furthermore from the replies to the questionnaire that no member state is currently considering creating a legal personality for AI-enabled robots in criminal matters[15], as criminal liability is unanimously based on intent or culpable negligence that can ultimately only be linked to a natural person, and this issue will therefore be excluded from the discussions, it being understood that research and debates are ongoing in this respect.

The CDPC Bureau decided at its meeting on 10 and 11 October 2019 "*to instruct the PC-CP to produce a study for the CDPC plenary in 2020 on the utility of drafting a standard-setting text in order to provide the necessary framework for the ever-increasing use of artificial intelligence by the prison and probation services in Europe*"[16]. The use of AI by the judicial system itself (risk assessment, profiling, predictive policing, facial identification, prisons and probation etc.) will not, therefore, be included in the discussions, as other bodies have been tasked with conducting an independent study on this subject.

Lastly, as in the preparation and dissemination of the questionnaire for member states on AI and criminal law[17], issues relating to cybersecurity will also be excluded from the analysis, to avoid any overlap with other Council of Europe activities and in particular those of the T-CY.

4.1.3. Definitions in the instrument

As things stand today, there is no universal definition of "Artificial Intelligence". In the CDPC's previous work, participants used the following definition of AI: a bundling of certain techniques – including mathematics logic, statistics, probability, computational neurobiology and computer science – with the goal of enabling a machine to imitate or even supersede the cognitive abilities of a human being.

Establishing a common working definition of AI is a precondition for any discussion and development of common standards in this regard in criminal matters. It will therefore be an essential discussion point when drawing up an international legal instrument on AI and criminal law, as will the correlative definitions of the terms "robot"[18] and "e-evidence"[19]. However, it will be sufficient to agree on working definitions within the scope of the instrument, i.e. for an instrument on driving automation it is possible to rely on technical standards (e.g. ISO norms).

4.2. Substantive criminal law: criminal liability of operators and providers of AI systems

A European denominator for substantive criminal law is the main issue at stake in the proposed Council of Europe legal instrument on AI and criminal law, namely the establishment of a common international framework for national substantive rules on criminal law and AI. Two basic issues, between many others, set out below could be discussed among the member states, the central and underlying ones being the liability approach regarding the possible risk arising from the employment of AI whether the current concept of negligence is sufficiently equipped to cover all blameworthy conduct or whether new special legislation is required.

Firstly, regarding the expected benefit of AI employment and the characteristics of human-robot-interaction, member statescould discuss whether they want to agree on a benchmark, a specific form of negligence and/or the extent of a harm caused or a level of automation put in place as trigger for a criminal investigation. If an instrument opts for a specific criminal proceeding, typical cases ought to be taken into account, where producers, users or other persons fail to take the necessary steps to control

[15] CDPC (2019)17, p. 5.
[16] CDPC-BU (2019) 4, point c, p. 2.
[17] CDPC (2019) 11, p. 3.
[18] For the purposes of its work, the Working Group used the following definition: a physically embodied artificially intelligent agent that can take actions that have effects on the physical world, but also a bot, i.e. an autonomous software program that can interact with other programmes or with a human user.
[19] For the purposes of its work, the Working Group used the following definition: data automatically generated during AI-driven human-robot co-operation that is offered as evidence in fact finding in a criminal trial.

risks emanating from robots, like flaws in the design or training of AI or when users disregard a system's instructions, for example by failing to take back control (reasonable time for regaining control of the machine) or, in the case of autonomously driving vehicles, by refusing to stop even though the system has detected signs of fatigue and suggested that the driver take a break. Possibly new forms of crimes must be defined, like speeding by faulty training of a speed assistant (that without plausible reason accelerates in a residential area) or dangerous interference with road trafficking by hacking.

Secondly, the existence and development of AI raises the question whether a new approach to criminal liability is needed, in cases where the offence is committed when the robot acts completely autonomous and/or the user has complied with all instructions, but the specifics of AI employment lead to harm. Taking AI as an acting counterpart into consideration when allocating criminal liability, at first blush, conflicts with basic principles of criminal law, which has been tailored for human action. However, if the human driver has vanished from the driver's seat entirely, the now passenger can no longer be held entirely responsible for accidents caused by the car. Some want to draw a parallel to corporate liability. However the parallel is not obvious, as corporations are legal persons in all member states, and thus liable under the law, but not competent to stand a criminal trial in all. Furthermore, in most countries corporate criminal liability connects to wrongdoing of a human representing the corporations. However, it seems obvious that the issue must be addressed, and the points to be discussed are: Could a producer of smart products or AI service provider be liable, as a question of principle, and if so, what level of negligence on its part and/or what degree of damage would be required in order for that provider to be criminally liable? Examples might include cases of criminal liability resulting from feeding in incorrect map data or sensor information, or designing an AI system that is dangerous. In this context, member states could also agree on a normative framework for the multiple forms of criminal liability that could be triggered by a combination of failures to exercise due care, resulting in an offence. This framework would then pave the way for nationwide reform of road traffic offences that currently focus on human action.

Lastly, member states could consider the nature and extent of the most appropriate criminal penalties for criminal liability involving an AI system, in the interests of legislative consistency at European level.

4.3. Procedural law and international co-operation: gathering evidence from AI systems

Resolving the issue of criminal liability is only meaningful if also inevitably related procedural problems are addressed, and this includes – among others – the issue of using the data generated through human-robot interaction as evidence.

Given the recent developments and applications of AI, the relevant evidence under criminal law is likely to be machine evidence, i.e. data generated by the robot, often taking part in a human robot interaction. This causes several problems:

First of all, availability of data could be a problem. Without further regulation, it would be consumer products, like cars, that generate data which is needed as evidence in a criminal trial. This data could be stored at the manufacturers of AI-enabled robots or with providers or with cloud service companies. It will then become an issue how to get access at home with a need for (harmonized) domestic procedural provisions and not to get access abroad with specific mutual legal assistance instruments. In this context consistency with the rules contained in other international instruments addressing this issue should be ensured, in particular with the Second additional protocol to the Budapest Convention, which is now being negotiated within Council of Europe (T-CY). Regarding international co-operation, classic problems connected to territoriality may be of particular concern, as, too, may be domestic laws protecting trade secrecy.

One solution could be a requirement to install a data registering box. The obligation to incorporate a recording device (sometimes called "black box", not to be confused with the opacity problem in robots

also referred to as black box-problem) when employing AI must be coordinated among states. It will be necessary to arrange for co-operation between member states to obtain data from the "black box" when it is no longer on the territory where the offence was committed, to establish an obligation for the manufacturer to disclose the necessary codes and information, the settings for machine learning systems, training data, etc., and to provide for oversight by judicial authorities.

Secondly, the issue of reliability and the vetting for trustworthiness in criminal and other proceedings arises, if consumer products, like cars, generate data which shall be used as evidence in a criminal trial. It could be necessary not only to regulate the collection, storage, encryption of and access to data but also consider that scope of trade secrecy privilege and other business protection. Traditional rules may not be designed to offer adequate means to meaningfully test the reliability and credibility of this new digital evidence nor sufficient protection for business interests nor the necessary instruments to tackle new problems, for instance the AI black box problem.

Thirdly, respect for fundamental rights in this area (right to private life and respect for the rights of the defence, in particular the right to silence and the right to question witnesses oneself conferred by Article 6 ECHR) must be addressed when developing a legal instrument on AI.

4.4. Preventive measures

In order to minimise the risks associated with the increasing use of AI systems, both in respect to criminal liability as well as with regard to evidentiary issues, the need for a general requirement for transparency as to the systems deployed and information on their operation, to be met by the private companies involved in the development and release of AI systems, ought to be discussed among member states. Similarly, the introduction of a general obligation to train users of AI systems could be the subject of negotiations among the member states, with driving schools, for example, being required to provide training in autonomous vehicles as part of the process of obtaining a driving licence.

In addition, member states may consider discussing their interest in (i) identifying instances where criminal prosecutions involving AI systems were brought before the domestic courts, the manner in which evidence was obtained in such cases and the legal standards according to which the decision was ultimately made, (ii) identifying any new national standards for AI and criminal liability, and (iii) ensuring that these national standards are consistent with the provisions of the future Council of Europe instrument. This independent national body would thus be in a position to alert the member state concerned to any problems relating to criminal liability and AI that would require the national or international normative framework to be adjusted or amended.

4.5. Protective measures

The establishment of a national or regional licensing mechanism for developers of AI systems could be encouraged by member states, enabling such licences to be suspended if a major criminal risk is identified, for example by the above-mentioned independent national body. This would be preferable to an outright ban on the use of AI systems on national territory, which would stifle the development of the technology.

4.6. Monitoring mechanisms

One final point to consider is the issue of monitoring the implementation of a future Council of Europe instrument. The latter would clearly benefit from a mechanism to ensure that it was being implemented properly. Even though most monitoring committees can produce only non-binding opinions (in contrast to the European Court of Human Rights), they nonetheless play a critical role in producing a collection of best practices that could serve as a model for others later on. Such high-level opinions could also be utilised by other Council of Europe bodies, such as the European Court of Human Rights, in future case law.

Conclusion

According to the plan and the steps to be taken as mapped out by the CDPC[20], it appears from the first output, namely the questionnaire compiled as part of the research project on national criminal law and the international legal framework regarding driving automation (or other AI deployment), that it is both highly desirable, and the wish of member states, that an international legal instrument be negotiated in the field of AI and criminal law, in line with the challenges and issues discussed above, so as to establish an international framework for developing specific national legislation. In effect, agreeing on common standards to clearly and properly allocate possible criminal responsibility and to clarify connected procedural issues as well as possible human rights implication needs to be a joint effort by public and private sector actors, so that the technology can develop successfully and in a way that respects the founding principles of civil society. It is the responsibility of member states to devise effective mechanisms to safeguard algorithmic accountability, working closely with those who develop and exercise digital power.

The CDPC could therefore move to the next stage of the plan (Concept Paper), namely output no. 2, consisting in the organisation of an international conference on common criminal law standards relating to harm caused by automated vehicles (or other AI deployment), providing a forum where member states, public and private sector actors can discuss developments in the field of AI, gaps in existing criminal law, legal solutions already in place or to be introduced via an international instrument, as well as output no. 3, consisting in the creation of an ad hoc drafting committee of experts for working on an instrument, whose form and content remain to be determined, establishing common international standards, first of all in criminal law relating to harm caused by automated vehicles (or other AI deployment).

[20] CDPC (2018) 14rev, p. 8.