Strasbourg, 10 March 2019 CDPC(2019)7

EUROPEAN COMMITTEE ON CRIME PROBLEMS

(CDPC)

Working Group of Experts on Artificial Intelligence and Criminal Law

WORKING PAPER II
1st meeting, Paris, 27 March 2019

Document prepared by Professor Sabine Gless, Special Rapporteur

A. Background: CDPC Initiative on AI & Criminal Justice

Following the Thematic Session on "Artificial Intelligence and criminal law" held on 28 November 2018, the members of the CDPC decided to establish a working group whose main task is to make an analysis of the impact of AI on criminal justice and present options for future CDPC work. The CDPC agreed that central to this (first) project is a common approach for topics central to CDPC work, aiming to prevent undesirable impacts caused by employment of AI and to prevent serious harm being caused by robotics.

The understanding is that the Council of Europe can play a key role in helping its member States to develop common legal standards which would provide an adequate, comprehensive and straightforward regulatory system, creating a workable equilibrium that recognises the many beneficial uses of such technologies while also guaranteeing accountability for any abuse and harmful consequences caused. Duly considering the *ultima ratio* of criminal regulation in this complex field, this project focuses only on situations where the level of harm, or the seriousness of the obligation breached, could or should entail criminal responsibility, or/and where the use of AI affects criminal justice systems directly.

For the purpose of our working group we will:

(1) start from the premise that ambient intelligent environments, or situations where pervasive computing is responding to humans' needs, result in increased human-robot co-operation in daily life. A poignant example of such co-operation is *automated driving, which has already foreshadowed the significant effects upon penal law and will be the primary example used for the purposes of this working group*.
However there are very similar issues in every field in which AI is deployed, be it robots that care for the elderly or IT-systems that carry out risk assessment in a prison.

(2) use the following definitions;

- **Artificial Intelligence (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.

- **Robot**, 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.

- **E-evidence**, data automatically generated during AI-driven human-robot co-operation that is offered as evidence in fact finding in a criminal trial.

- **Driving automation**, the deployment of AI to gradually replace the human driver with driving assistants that (temporarily) take over driver's tasks. Currently, the industry distinguishes between five[1] levels of automated driving.

[1] Norm SAE J3016_201401 <https://www.sae.org/standards/content/j3016_201401>.

B. AI & Criminal Justice

In November 2018, during the Thematic Session, the CDPC identified the following areas of interest where AI has an impact on criminal justice and that could affect the general principles of criminal law, for example due process, fairness, the basic concepts of cross border co-operation:

- Substantive Criminal law: in particular the risk of a responsibility gap;
- Procedural Law: in particular the problem of eEvidence;
- Mutual Legal Assistance: in particular dual criminality and transfer of evidence across borders;
- Penitentiary law and law enforcement law: in particular risk assessment.

The following *tour d'horizon* provides a short introduction to each field before we turn to the practical issue of drafting a questionnaire. It is understood however that the *primary topic is challenges to substantive criminal law* when the human actor disappears (if we go for example from coach to car to driverless), but the need to allocate responsibility remains if harm is caused through the human-robot co-operation. Substantive criminal law must address the issues of agency, a negligent collaboration of various actors that results in harm, and socially accepted risks. When an incident happens, the relevant evidence is likely to be machine evidence, i.e. data generated by the robot part of the co-operative relationship, which also poses new challenges for mutual legal assistance. The data must be retrieved and meaningfully integrated into fact finding for criminal proceedings which may in turn affect defence rights as well as raise questions of territoriality principle depending where the information is stored. All these problems are paradigmatic when humans and machines co-operate, which can also be the case to serve criminal justice directly. For instance when AI do risk assessments for early release from prison, and humans accept such decisions, questions of responsibility, credibility and transferability across borders also arise.

I. Substantive Criminal Law

1. Is there a responsibility gap?

Will there be a responsibility gap (when the human actor disappears)?

When human actors co-operate with robots and human action is subsequently replaced by robot acts *(i.e. driving automation systems gradually replacing the driver)*, the question of who is responsible for damage caused by the robot conduct *(i.e. driver, producer, provider)* arises.

Ultimately the goal is to ensure adequate accountability for robots' acts. For this reason government regulation on liability issues is an important component. The scope and content of such provisions are to be determined and could cover (specific) situations of human-robot-co-operation, such as drivers' liability for automated driving which is addressed in a German provision[2] or general rules of liability (for instance Art. 12 of *Convention* on *Cybercrime, ETS*

[2] § 1a Strassenverkehrsgesetz /Federal Law on Road Transport adopted 2017 www.gesetze-im-internet.de/stvg:
"Driving an automated vehicle is legal as long as the automated driving systems are used consistent with the authorization …"

No.185 obligating States to ensure corporate liability). Such regulation could also entail compliance rules for companies in the AI industry to ensure proper legal representation and internal governance structures.

2. How to define the socially permissible risk?

Modern life in general and motorised traffic in particular brings with it a risk to life and limb. Nevertheless it is legal to drive a car as society accepts certain risks when it comes to road traffic. Car producers and other experts consider that automated driving (based on a human-robot co-operation) will be safer in many situations than human driving. As the crucial question in the development of automated driving might concern what kind of risk respective societies are willing to accept, it is important to understand whether all States share the notion of a socially permissible risk. This can also be an issue for MLA (see infra xxx) and the question of dual criminality. If States could agree on a common definition, the potential for conflicts could be reduced.

3. Can responsibility be allocated (among several providers for AI Systems)?

Robots often rely on many inputs and services to be able to function. If damage is caused it can be difficult to determine who is responsible for a certain data intake and output. These problems are addressed at a technological level, like the black box-problem, but must also be addressed with a regulatory approach. The content of such regulation could entail liability for the supply chain,[3] definition of risk spheres, etc.

II. Criminal Procedure

1. Challenges of Machine Evidence

Where robot-human co-operation causes harm and a human driver faces negligence charges, the relevant evidence presented against him or her is likely to be “machine evidence” or data generated by the robot during the activity in question. The question arises of how such data can be retrieved and read out (see also MLA infra xxx) and, if presented as evidence, how it can be tested for credibility. Fact-finding procedure, and especially the assessment of the reliability of evidence, is surprisingly human-centred. Here, research suggests that the adversarial system offers, in principle, valuable components to develop ways of efficiently testing the reliability of machine evidence. But, up to now, the inquisitorial system does not.

2. Defense Rights (Art. 6 paragraph 3 ECHR, especially the right to “examine a witness”)?

The use of machine evidence might weaken defence rights. If, for instance, a lethal incident occurs during a car journey partly handled by a human driver and partly by a driving automation, and only the human driver is to be prosecuted, the question arises whether the driver will have adequate facilities for the preparation of a meaningful defense with regard to the robot-

[3] as the entities co-operating to provide AI probably best know how to organise accountability and transparency along the supply chain (origins and use of training data, test data, models, application programme interfaces (APIs) etc.)

generated data presented against him or her. The right for a person charged with a criminal offence to examine witnesses him/herself, granted by Article 6, paragraph 3, lit. d of the ECHR could be a starting point to press for access to the source code, the machine learning parameter, training data, etc.

3. Monitoring Code & Industry (Open Source/Trade Secret Privilege/Whistle Blowers)

Possibly a meaningful defence (examination of machine evidence) requires independent monitoring and transparency of AI systems from the beginning. Only if third party experts were able to audit and publish information about key systems independently from a criminal trial and AI infrastructures were understood from set up to training to deployment, then defence strategies could be developed. This raises the issues of open source, trade secret privilege and protection for whistle blowers, i. e. conscientious insiders who step forward to disclose valuable information at a criminal trial.

4. Intrusive Investigations and Human Rights

Law enforcement agents, prosecution services and other authorities may use AI systems, including cars, in ways that affect the validity of criminal investigations. The huge developments in AI face recognition programmes could be valuable in the identification of alleged perpetrators or convicts on the run. However, the use of such instruments may infringe on Article 8 of the ECHR and possibly needs stringent regulation to protect privacy. AI is used for profiling, including programmes for predictive policing which raise issues of racial discrimination or the validity of the presumption of innocence.

III. Mutual Legal Assistance MLA (and Infrastructural Concerns)

1. Dual Criminality

The requirement of dual criminality is a traditional principle of MLA, according to which a country will not assist in the prosecution of conduct it does not deem criminal. In the EU the application of this principle is narrowed down, but in the CoE context it generally applies. If some States allow for automated driving and others not, but cars cross borders, problems could arise from drivers using technology they are only allowed to use in a particular country. In theory, this is not a problem as traffic law is territorially applied, with or without driving automation. But in practice, an approved vehicle for automated driving that must be driven by hand in a particular country could give rise to problems.

2. Access to Data/Evidence across Borders

Access to data involving AI instruments, including driving automation, often requires cross-border activity, as servers may be in a foreign jurisdiction or data stored in a cloud provided by a foreign private company. The Convention on Cybercrime ETS 185 (CCC) gives cross-border access to data with minimum requirements on investigative measures, including production orders (Art. 18 CCC) and preservation orders (Art. 16 & 17 CCC). But it is not entirely clear whether the means provided by the CCC are sufficient or whether an update is required.

3. MLA & Private Stakeholders

Today, data often is stored with a private Cloud Service Provider (CPS), possibly in a third country. In certain criminal investigations territorial rules no longer seem to be the gatekeepers of access to information; instead it is now the CSPs. For MLA purposes this must be kept in mind.

IV. Penitentiary Law, Policing / Risk Assessment

The employment of AI could also affect penitentiary law or policing. In various areas programmes that profile individuals in a risk assessment, e.g. inmates for recidivism risks to decide upon early release from prison, are used. Tools for effective recognition in prisons may be used to decide upon living conditions provided to different inmates.

One could envisage to use such tools in cars, for instance to detect road rage or a tendency to break the speed limit.

C. Further action

1. Work Stages

The working group should discuss, but not take a final decision on the overall goal:

(a) with regard to substance (e.g. how many of the areas depicted under B. I.-IV. shall be covered?);

(b) with regard to form (what form of CDPC activity in the field is envisaged, (i.) updating a Convention, e.g. the Cybercrime Convention with a protocol; (ii.) drafting & adopting a new Convention "AI and Criminal Law Convention"; (iii.) a recommendation or other soft law instrument?

2. Questionnaire

The working group is to draft a questionnaire.

3. Work Plan

27 March 2019	1st meeting of the working group (Paris) to prepare the questionnaire mapping criminal law and criminal procedure in member States
April/May 2019	Send the questionnaire to all CDPC delegations (member States)
September 2019	Deadline for member States to send their replies to the questionnaire
September/October 2019	Preparation of an analysis of the replies to the questionnaire received by member States
October 2019	2nd meeting of the working group (Paris) to discuss the results achieved to be presented at the CDPC plenary meeting in December
3-6 December 2019	Presentation of the results achieved by the working group at the CDPC plenary meeting (Strasbourg)
January/February 2020	3rd meeting of the working group (Paris). Considerations on findings of CDPC and start working on CoE / CDPC instrument
June 2020	International Conference on common criminal law standards relating to harm caused by automated vehicles (or other AI deployment)
September/October 2020	4th meeting of the working group (Paris) to finalise the draft instrument for presentation to the CDPC