RESOLUTION ON

TRADITIONAL CRIMINAL LAW CATEGORIES AND AI

By Lorenzo Picotti[*]

Preamble

Considering that

- Artificial Intelligence (AI) systems with different degrees of autonomy support and replace many human activities;
- AI systems might represent a real benefit for society as a whole and for the law enforcement community, specifically when it comes to the investigation of criminal offences;
- AI systems are becoming increasingly autonomous, and their functioning may be unpredictable to those who design, program, produce, distribute, and use them;

Observing that

- AI systems' areas of application of are considerably wide, and illicit facts related to their implementation might harm various interests, legal goods, and fundamental rights;
- AI systems may also play an increasingly role in the perpetration of criminal acts as 'tools' to commit criminal offences, and they might become the facilitating factor of the emergence of new criminal acts;

Paying particular attention to

- the increasing full or partial delegation of decisions to AI systems in different areas of activity, which raises the question of natural or legal persons' liability for the harms caused by the autonomous functioning of such systems;
- the autonomy of AI systems, and the possibility debated within academia of considering them perpetrators of a crime;

Bearing in mind

- the importance of appropriate reactions criminal law is required to provide to prevent and punish offences committed by, through, or against AI systems;
- the seriousness of the harms and of the risks related to AI applications;
- the fundamentals principles that must be guaranteed in establishing and applying criminal sanctions (including punitive sanctions in a broader sense, which could be applied to legal persons), such as the principle of legality and the principle of culpability, which is a necessary expression of the personality of penal responsibility;

[*] Full Professor of Criminal Law, University of Verona – lorenzo.picotti@univr.it.

- that traditional criminal law categories and models of criminal responsibility need to be considered and, if necessary, adapted to emerging requirements for protection;

Taking into account

- the 'Ethics Guidelines for Trustworthy AI' presented to the European Commission on 8 April 2019 by the High-Level Expert Group and other significant recommendations by other international bodies (for instance the 'Feasibility study on a future council of Europe instrument on Artificial Intelligence and Criminal Law' of the European committee on crime problems of the Council of Europe, 4 September 2020);
- the proposed European Regulation on Artificial Intelligence (so-called AI Act); the work of the Committee on Artificial Intelligence of the Council of Europe; the United Nations Activities on Artificial Intelligence;
- the recommendations of the XIV International Congress (Vienna, 1989) on the legal and practical problems posed by the difference between criminal law and administrative penal law; those of the XVIII International Congress (Istanbul, 2009) on the incrimination of preparation and participation in a crime; and those of the XIX International Congress (Rio de Janeiro, 2014) on Information Society and Penal Law;

The participants of the International Colloquium of Section I of the XXI International Congress of Penal Law (Criminal Law: general part): 'Traditional Criminal Law Categories and AI: Crisis or Palingenesis?' have adopted what follows.

Recommendations

1 On the concept of Artificial Intelligence and the attribution of legal personality to AI systems with different degrees of autonomy

1. The concept of AI encompasses multiple algorithmic and robotic systems that interact with the environment, developed using several techniques (such as machine learning), for pursuing human-defined objectives. It is therefore not desirable to provide a general definition of AI for criminal-law purposes.

2. However, since AI systems can be harmful in several fields (e.g., self-driving vehicles, robotic systems in medicine, AI trading systems or logistics management), the protection of legal goods and fundamental rights under criminal law should take into account the specific features of the various AI systems with different degrees of autonomy, as well as the legal definitions provided by extra-criminal-law sources in each specific sector.

3. As it stands, there is no normative ground nor consistency relating to the functions of criminal punishment in recognizing legal personhood for AI systems with different degrees of autonomy.

4. On the one hand, there is an ontological distinction from human agents. AI systems lack the consciousness in choosing and evaluating the possible solutions to a problem or

dilemma, considering also the context of social and ethical relationships and opportunities, with the necessary flexibility and capacity to adapt to even contingent or supervening situations and conditions.

5. On the other hand, punitive sanctions applied to such technological systems and agents would not respond to the purposes and functions of criminal punishment because the effect of the threat of the penalty and its application would be emptied by the absence of self-awareness of their own existence in the past, present, and future, and, above all, by the absence of voluntary self-determination, so that, even excluding the retributive function, not even those of special and general prevention would be feasible.

2 On the need for extra-criminal regulation, standards, and obligations

6. To prevent and reduce AI-related harms before or, at least, in parallel with criminal-law reforms, it would be necessary for international, regional, and national legislators and competent authorities to fully define the regulation of the several sectors in which AI systems are implemented (such as those mentioned above of self-driving vehicles, health and surgical robots, autonomous weapons, etc.). The technical standards, structural characteristics, and operating conditions of AI systems and their components should be regulated.

7. Such regulations, which must operate from the design, production, distribution, and sales phases through the actual use of AI systems should also provide for concrete requirements concerning adaptation in the event of red flags or warning signals as pre-conditions for addressing AI-related harms through punitive law.

8. These regulations might provide for injunctional procedures, such as those already provided for in areas of complex risks (e.g., health and safety in the workplace and environmental protection), the violation or non-compliance of which may be sanctioned according to the *ultima ratio* principle.

3 On the need for criminal protection of legal goods

9. It is necessary to recognize the essential importance of a reasonable and proportionate intervention of criminal law in a broad sense in preventing and punishing harms and dangers AI systems might cause to interests, legal goods, and fundamental rights, which, given the same facts, might constitute a criminal offence according to the traditional categories of criminal law if realized by natural and legal persons. Therefore, they cannot go unpunished simply because they are carried out by, through, or against the aforementioned systems.

10. It is necessary to identify and define specific models for attributing liability to the persons (both natural and legal persons) who are 'behind' AI systems (i.e., the actors in the different phases of the life cycle of AI systems, such as designers, providers, importers, distributors, users, etc.), starting with the owners and those who decide on their concrete use, based on their interest and their benefit, and who must therefore be held legally liable, including from a 'punitive' – not only a criminal – perspective.

11. The responsibility of the persons described in the previous point does not exclude that of other persons (either natural or legal persons) who contribute to the causal chain of the harm, from the designers, programmers, producers, sellers, and distributors to the end-users of the systems themselves.

12. In particular, a distinction must be made between:

a. AI systems used in illicit activities. In this area, there will mainly be malicious conducts, which pose fewer problems in terms of attribution of criminal liability, given that AI systems are conceptually no different from other instruments and means of committing a crime.

Two issues, however, should be addressed:

a.1. in case of deviant results of the functioning of the system from the intended illicit activity, the traditional principles of *aberratio ictus* and *aberratio delicti* must be applied. The mere material diversity of the harmed object must not represent an excuse if its characteristics are not relevant for the configuration of the criminal offence (e.g., killing one person instead of another is not relevant for the realization of the crime of murder, when it is intended by the agent). Instead, it is preferrable to base criminal liability for a crime other than the one intended on the possibility of concretely foreseeing such a different development of the action put in place by the AI system by applying the principles of negligence-based liability (as set out in paragraph IV below);

a.2. since AI systems can be used for particularly harmful or dangerous conducts, they can amplify and aggravate the harm caused (as happen with ICT). Indeed, the consequences can be very distant from the actions that gave origin to them, making it more difficult to intervene *post factum* to prevent or at least to end or reduce the harmful events. Therefore, the incrimination, as autonomous preparatory offences, of illegally designing, programming, producing, distributing, selling, and purchasing 'malicious' algorithms, software, and AI systems should be considered. This criminal policy should be limited to AI systems that pose high risks to certain significant legal goods (such as life, body, or liberty of other human beings) and only in cases of clear, actual, present danger (on the conditions required in incriminating preparatory acts, see the resolution of the Section I of the XVIII AIDP Congress in Istanbul, 2009).

b. AI systems used in lawful activities. This case raises the most delicate issues in the area of 'permitted risk', which should be delimitated through the hoped-for regulation of specific security obligations and precautionary rules to be applied to the activities of design, development, production, distribution, and sale as well as use of AI systems. The adjustment of the models of criminal liability in this area must address the friction that can be created between forms of responsibility for negligent behaviour and the technical features of AI systems, namely: (1) their autonomy; (2) the concrete unpredictability of their decisions and functioning; (3) the opacity of their regulatory mechanisms; (4) the complexity of their programming, development, production, updating, and maintenance process.

4 On the adaptation of the models of attribution of liability to the features of AI systems, specifically to their degree of autonomy

13. First of all, a distinction must be made, according to the recognized graduated automation and autonomy of AI applications in several areas, between the different levels of decision-making and operational autonomy of AI systems, which go from those in which the 'automatic' functioning allows the human agent to have significant control over the system to those that are truly 'autonomous', in which human intervention can only be distant, in time and in space, from the functioning of the AI system, which 'decides' based on information collected and on algorithms that adapt to its experience, so that there is a structural margin of unpredictability of the concrete outcomes.

14. In relation to the different types of AI systems, the definition of specific rules and standards of functioning, as foreshadowed in the proposal for a European regulation on Artificial Intelligence, is of fundamental importance (see paragraph II).

15. The most pressing need for adaptation of the traditional categories of criminal law concerns the area of AI systems with a greater degree of autonomy towards which current technological development and experimentation in many fields is tending, meaning they will undoubtedly be even more important in the near future.

16. Under this perspective, the field of corporate punitive liability might provide a useful reference, as night [?] the fields of product liability and liability for the protection of health and safety in the workplace.

17. In these legally regulated fields, often harmonized at a European level, existing principles might be extended, with the necessary adaptations, to AI-related crime regulation. Such regulated fields require the preventive assessment of the risks inherent in the specific activities performed, which have margins of permitted risk and correlated obligations of risk-prevention and containment (see paragraph II), with specific regard to the sources of dangers and harms.

18. Duties to act, especially in the case of red flags, are imposed to the relevant categories of persons (human beings) operating according to their respective competences, i.e., users and persons having the position of guarantee. They must promptly adapt the regulatory and security measures appropriate to their activity, to the point of stopping it, if necessary.

19. From these recognized principles, the following recommendations can be elaborated to structure criminal liability for AI-related harms:

a. Criminal responsibility. of natural persons. This must be based on the identification of personal positions of guarantee in relation to the competences and functions performed in using AI systems. Firstly, the actors participating in the different phases of use of AI systems, up to the end-users, will have to be considered. Secondly, the positions of top and middle management and compliance officers in complex organizations should be considered. In each case, the formalization of positive obligations of a technical, organizational, and control nature shall be addressed.

Criminal liability for negligent behaviours must comply with the general principles of criminal law, namely the principle of personal culpability, since the objective connection between the causal contribution of the human agent and the commission of the offence by the AI system does not suffice, given that the foreseeability and evitability of the illicit fact are also necessary. Particularly, criminal responsibility for negligence – for not having acted differently from what would have been possible – must be correlated not so much to the specific and concrete event or fact that occurred as to the scheme of 'organizational fault', referring to the way the artificial agent is structured and operates. The assessment of the risks arising from the AI system's activities must also include the awareness of outcomes that are concretely 'unforeseeable' in individual cases, which is the basis of the obligation to prepare adequate and always up-to-date surveillance and containment measures for which the natural person in charge remains responsible, being accountable (accountability), as the owner or top representative of the organization using the AI system in its own interest or to its own advantage.

b. Punitive liability of legal persons. Considering that a large part of AI systems is produced or used by legal persons, it is necessary to hold them accountable for the offences committed by, through, or against such systems.

In this respect, assuming that precise public standards of conduct and compliance are to be introduced (cf. paragraph II), punishment of the legal person, proportionate to the offences committed by, through, or against AI systems and to the degree of fault of the organization, might be related to a model of liability based on organizational fault. Such a model of liability leads to imputing responsibility subjectively, as the object of culpable reprehensibility, to the legal person in the event of offences caused by the lack, deficiency, or inadequacy of organizational and prevention measures to be implemented and updated on the basis of assessment of the specific risks deriving from the activities entrusted to and, in any case, carried out by AI systems in their interest or to their advantage. Strict liability models should be avoided.

A new model of corporate autonomous punitive liability, not based on the liability of the individual natural person, should be promoted, since the legal person can be held liable even if the natural person who realized the harm is not individually punishable due to particular conditions or circumstances or if he/she is not specifically identified. Indeed, it is enough to ascertain the commission of an objectively typical and unlawful act in the interest or to the advantage of the organization.

In those national legal systems that base corporate liability to a closed list of offences, the extension of such list to the criminal offences which can be committed through, by, or against AI systems is recommended.

5 On preventive measures and punitive sanctions applicable to natural and legal persons 'behind' Artificial Intelligence systems

20. The sanctions applicable according to the various legal systems to natural persons, including imprisonment, and to legal persons, possibly of an administrative nature but

in any case of a punitive nature, including fines and suspension of the activity by which the offence was committed, should in principle correspond to those applied for the type of offence realized, in accordance with the principles of each legal systems, namely the principles of proportionality and individualization of sanctions. When it comes to legal persons, these sanctions, in addition to pecuniary measures, should also include the injunction to modify the corporation's compliance and internal control system, as well as the possibility of ordering a period of public monitoring of the corporation to ensure that it complies with the imposed standards.

21. Given the seriousness of AI-related harms, the adoption of preventive measures for their effectiveness in averting or mitigating the harmful consequences of AI systems (as, *inter alia*, seizure, confiscation, judicial monitoring, interdictory measures) is recommended.

22. The important role of non-pecuniary penalties, such as disqualification from exercising specific activities and confiscation, should be emphasized. Specifically, confiscation allows direct action to be taken against the AI system with or by which the offence was committed, without the need to recognize it as a legal entity or as having criminal capacity (see paragraph I above).

6 Complementary enforcement systems

23. Given the problematic and foreseeable difficulty of implementing an effective criminal-liability system for natural persons and legal entities 'behind' AI systems for offences committed by, through, or against them, the adoption of complementary enforcement systems is recommended.

24. Such enforcement system might include administrative authorizations and certifications, as well as civil remedies.

25. Other alternatives to criminal prosecution and/or sanctioning might include models of compliance, restorative justice interventions, and agreements with victims and competent public authorities.