DIRITTO PENALE
E UOMO
Criminal Law and Human Condition

Riflessione

Artificial Intelligence in Healthcare: Risk Assessment and Criminal Law

Intelligenza artificiale in ambito sanitario: valutazione del rischio e diritto penale

di Federico Carmelo La Vattiata

Contributo originariamente pubblicato all'interno del volume collettaneo *Special Collection on Artificial Intelligence*, edito dal United Nations Interregional Crime and Justice Research Institute (UNICRI), 2020, pp. 4 ss.

***

Il presente contributo, del quale è stata autorizzata la ripubblicazione all'interno di questa Rivista, è parte integrante del volume collettaneo *Special Collection on Artificial Intelligence*, recentemente pubblicato dall'Istituto di ricerca sul crimine e la giustizia delle Nazioni Unite (UNICRI). UNICRI ha in tal modo voluto realizzare un'opera di carattere interdisciplinare su un argomento che, verosimilmente, impegnerà i giuristi nei prossimi anni. Occorrerà individuare, infatti, delicati punti di equilibrio tra diversi interessi, certamente non tutti aventi la medesima importanza avuto riguardo alle scale valoriali consacrate nelle Carte costituzionali e nelle Carte internazionali in tema di diritti umani, ma comunque meritevoli di protezione (sia pure di variabile intensità) da parte del diritto.

A ben vedere, la *Special Collection* rappresenta solo l'ultimo "tassello" del più ampio programma di ricerca dell'Istituto (in particolare dal suo *Centre for Artificial Intelligence and Robotics*) in materia di rapporti tra l'intelligenza artificiale (IA o AI) e il settore della giustizia penale. L'impegno profuso è testimoniato dalla pubblicazione di diversi lavori (oltre a quello in parola), tra i quali vanno rammentati: *Artificial Intelligence: an overview on state initiative* (luglio 2019); *Artificial Intelligence and Robotics for Law Enforcement* (marzo 2019); nonché *Towards Responsible Artificial Intelligence Innovation* (maggio 2020) – queste due ultime opere prodotte in collaborazione con INTERPOL –.

L'articolo qui ripubblicato, in particolare, analizza il tema della gestione dei rischi legati alle attività di produzione e impiego di sistemi di IA, con particolare riguardo alle applicazioni in ambito medico (segnatamente i *Software as a Medical Device* o SaMD). L'intento è quello di definire le questioni rilevanti sotto il profilo penalistico. A questo riguardo, viene indagata la tenuta delle categorie tradizionali del diritto penale di evento, a fronte di una così dirompente innovazione tecnico-scientifica. L'approccio adottato si contraddistingue per due elementi: l'interdisciplinarietà e il metodo della comparazione. Da quest'ultimo punto di vista, lo studio fa riferimento a tre tradizioni giuridiche differenti (quella di *common law*, quella tedesca e quella italiana), allo scopo di cogliere gli elementi ad esse comuni, senza con ciò ignorare i tratti differenziali. Infine, viene proposto un modello alternativo di *AI-risk assessment*, basato su una integrazione degli strumenti propri del diritto penale, del diritto civile e del diritto amministrativo. Invero, alla luce del principio del diritto penale come *extrema ratio*, una corretta calibratura delle sfere di rilevanza di tali settori dell'ordinamento garantirebbe la contemporanea tutela degli interessi, da un lato, dell'imputato/indagato a un processo equo e, dall'altro, delle vittime a una soddisfazione per i danni subiti.

***

Abstract. The advances in the field of artificial intelligence (AI) are changing the nature of medical care. They involve both the sectors of diagnostics and therapeutics. Medical literature has widely analysed the advantages and the risks of AI. Researchers have found that early diagnoses are essential in order to avert the decline of patients' health status. This can be achieved through improving the analysis procedures on healthcare data by means of AI techniques. However, in order to guarantee the security of AI medical devices, their clinical evaluation is crucial.
This article aims at conceptualising and solving the questions related to the application of AI in healthcare from the point of view of criminal law. The traditional criminal law categories will be investigated, so as to understand whether it is possible to consider deaths and injuries occurring in the context of medical care as criminal offences to prevent and prosecute, when AI techniques are used. The study will be carried out in a comparative perspective. In conclusion, this will allow to propose a new AI-risk assessment paradigm, based on the integration of criminal law, civil law, and administrative law measures, so as to guarantee an equilibrium between the fundamental rights of the accused (a fair trial) and of the victims (a compensation for damages they have suffered).

Abstract. I progressi nel campo dell'intelligenza artificiale (IA) stanno modificando l'essenza delle cure mediche. Essi interessano sia il versante della diagnostica sia quello della terapeutica. La letteratura medica si è diffusamente soffermata sull'analisi dei vantaggi e dei rischi connessi all'IA. I ricercatori hanno riscontrato come le diagnosi precoci risultino essenziali per evitare il peggioramento delle condizioni di salute dei pazienti. Questo risultato può essere conseguito perfezionando le procedure di analisi dei dati sanitari tramite il ricorso a tecniche di IA. Tuttavia, per garantire la sicurezza dei dispositivi medici di IA, è fondamentale la loro valutazione clinica.
Il presente contributo mira a inquadrare e risolvere le questioni relative all'applicazione dell'IA in ambito sanitario dal punto di vista del diritto penale. Verranno indagate le tradizionali categorie del diritto penale, al fine di comprendere se sia possibile ricondurre i decessi e le lesioni occorsi in ragione dell'utilizzo di applicazioni mediche di IA a fatti-reato da prevenire e perseguire. La disamina sarà condotta secondo una prospettiva comparativa. All'esito dell'analisi, sarà possibile proporre un nuovo paradigma di valutazione del rischio connesso all'IA, basato sul ricorso integrato a misure di

diritto penale, diritto civile e diritto amministrativo, allo scopo di assicurare un equilibrio tra i diritti fondamentali dell'imputato (garanzia di un processo equo) e delle vittime (diritto al risarcimento dei danni subiti).

1. Introduction.

The advances in the field of artificial intelligence are changing the nature of medical care. AI refers to «systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world [...] or AI can be embedded in hardware devices»[1]. In other words, an **intelligent system** consists of a set of algorithms that can use data, to solve (more or less complex) problems in different contexts.

One should clarify a fundamental technique in the field of AI, i.e., **machine learning** (ML). The latter is based on complex mathematical techniques, since the related knowledge-representations involve the theory of probability and the statistics. The **artificial neural nets** (ANNs) are crucial elements. They are neural computational systems that are inspired by the functioning of the human brain, i.e., **biological neural nets** (BNNs). In particular, there are two similarities between them. Firstly, the building blocks of both nets are highly interconnected computational "tools". Secondly, ANNs consist in computing networks that are distributed in parallel and function like the varying synaptic strengths of the biological neurons: there are many **input signals** to neurons, and the impact of each input is affected by the **weight** given to it, namely the adaptive coefficient within the network that determine the intensity of the input signal. In conclusion, «the output signal of a neuron is produced by the summation block, corresponding roughly to the biological cell body, which adds all of the weighted inputs algebraically»[2].

There are three kinds of ML:

[1] EU Commission, *Artificial intelligence for Europe*, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2, April 25, 2018.
[2] Y.-S. Park, S. Lek, *Artificial Neural Networks: Multilayer Perceptron for Ecological Modelling*, in *Developments in Environmental Modelling*, 28, 2016, p. 124.

a) the **supervised learning**, where the programmer **trains** the system by defining a set of expected results in relation to a certain in-put range, and by constantly evaluating the achievements of the objectives. The system then formulates a hypothesis. Every time it makes a mistake, the hypothesis is reviewed;
b) the **unsupervised learning**, where the user provides for neither expected results nor error-reports;
c) the **reinforcement learning**, where the system is led by a sort of *reward-punishment* mechanism, i.e., feedback messages about what has been done well or badly. In complex situations, the success or the failure of the system is reported after many decisions, and a sort of procedure for **assignment of credits** identifies the decisions that likely lead to success.

Another important technique is the so-called **deep learning** (DL), a subset of ML: it is a technique that allows the machines to better identify patterns through a more complicated use of neural nets. Because of this characteristic, it is possible to recognise data's patterns at various hierarchical levels. In other words, DL can identify a multilevel representation of knowledge[3].

In brief, the ML algorithms use statistics and mathematics to find a pattern and correlations in big data[4]. The more a system processes data, the more it improves. Hence, the **detention** of big data is important. Very few persons and entities hold this new kind of **wealth**. This involves the emersion of new power, which is particularly difficult to control and limit by virtue of the traditional means. For this reason, new **checks and balances** are needed, so as to balance two interests: *a)* to promote the advantageous developments of AI; and *b)* to prevent abuses that can cause threats to individual rights[5].

In general, there are four classification criteria of risks and opportunities resulting from the use of AI. When the AI is well used, it allows the human **self-regulation** and **agency**, as well as the improvement of societal **potential** and **cohesion**. Instead, when the AI is misused or overused, it reduces the **competences** of humans, it removes their **responsibilities**, and it undervalues their skills of **self-control** and **self-determination**[6].

2. Ethics Guidelines for Trustworthy Artificial Intelligence.

In 2018 the European Commission set up an independent High-Level Expert Group on Artificial Intelligence (AI HLEG) that published a document, entitled "Ethics Guidelines for Trustworthy Artificial Intelligence", with the aim to promote trustworthy AI.

---

[3] M.A. Boden, *Artificial Intelligence. A Very Short Introduction*, Oxford University Press, 2018, (Italian trans. *L'intelligenza artificiale*, Il Mulino, 2019), p. 46.
[4] The definition of "big data" is: very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods.
See https://dictionary.cambridge.org/it/dizionario/inglese/big-data.
[5] C. Casonato, *Potenzialità e sfide dell'intelligenza artificiale*, in *BioLaw Journal*, 1, 2019, p. 178.
[6] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, *AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and machines*, 28, 4, 2018, pp. 689 ff.

According to the document, «trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems»[7]. Trustworthy AI should be: *a)* lawful; *b)* ethical; and *c)* robust.

The document provides a framework for achieving trustworthy AI based on fundamental rights, that are enshrined in the Charter of Fundamental Rights of the European Union, and relevant international human rights law, such as the European Convention on Human Rights. Furthermore, the document sets three series of key guidance. First, it «identifies the ethical principles and their correlated values that must be respected in the development, deployment and use of AI systems»: *a)* respect for human autonomy, prevention of harm, fairness and explicability; *b)* the need for attention to situations involving more vulnerable groups, and to situations which are characterised by asymmetries of power or information; and *c)* acknowledgment of the risks that AI involves (including impacts which may be difficult to anticipate, identify or measure), and adoption of adequate measures to mitigate them. Secondly, guidance on how trustworthy AI can be realised is provided, and for this purpose seven key requirements that AI systems should meet are listed: *i)* human agency and oversight; *ii)* technical robustness and safety; *iii)* privacy and data governance; *iv)* transparency; *v)* diversity; non-discrimination and fairness; *vi)* environmental and societal well-being; and *vii)* accountability. Finally, the document provides an assessment list that will need to be tailored to the specific use case of the AI system.

As we will see in the following paragraph, AI can be widely applied in medicine, and its potential contributions to this field seem limitless. However, ethical challenges can arise, given that intelligent systems have a tremendous capability to threaten patients' preferences, safety and privacy. Benefits and risks due to AI application in healthcare need to be balanced carefully[8]. Furthermore, criminal law issues can arise with respect to the area of medical malpractice, such as black-box problems concerning the etiological link between the use of AI and damages, as well as difficulties in identifying precautionary rules whose culpable violation justifies convictions for gross negligence or recklessness.

3. AI in Healthcare.

Experts started to debate the topic of difficult decisions in complex clinical situations assisted by computers in 1959[9].

[7] EU High-Level Expert Group on AI, *Ethic Guidelines for Trustworthy AI*, 4, April 8, 2019.
[8] M.J. Rigby, *Ethical Dimensions of Using Artificial Intelligence in Health Care*, in *AMA J Ethics*, 21, 2, 2019, pp. 121 ff.
[9] R.S. Ledley, L.B. Lusted, *Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason*, in *Science*, 130, 3366, July 3, 1959; E.H. Shortliffe, M.J. Sepúlveda, *Clinical Decision Support in the Era of Artificial Intelligence*, in *JAMA*, 320, 21, December 4, 2018, p. 2199.

AI in healthcare involves both the sectors of diagnostics and therapeutics[10]. Applications are diverse, from mobile apps that make a diagnosis to surgical robots[11]. The advertising hyperbole of AI medical devices, however, «has led to skepticism and misunderstanding of what is and is not possible» with ML[12]. One should investigate the reasons for the doubts of credibility that affect the adoption of clinical decision support systems (CDSS). There are complexities that limit the ability to move ahead quickly. They reflect the ones of clinical practice: *a)* black boxes are unacceptable, since CDSS require transparency; *b)* time is a scarce resource, given that CDSS should be «efficient in terms of time requirements and must blend into the workflow of the busy clinical environment»; *c)* «complexity and lack of usability thwart use», as CDSS «should be intuitive and simple to learn and use»; *d)* «relevance and insight are essential», in fact «CDSS should reflect an understanding of the pertinent domain and the kinds of questions with which clinicians are likely to want assistance»; *e)* delivery of knowledge and information must be respectful, and CDSS «should offer advice in a way that recognizes the expertise of the user, making it clear that it is designed to inform and assist but not to replace a clinician»; and finally *f)* scientific foundation must be strong, as CDSS «should have rigorous, peer-reviewed scientific evidence establishing its safety, validity, reproducibility, usability, and reliability»[13].

Although human physicians cannot be replaced by AI systems in the foreseeable future, AI could play a key-role in assisting physicians to make better clinical decisions. In some cases, intelligent systems could even replace human judgement in certain fields of healthcare.

As a matter of fact, a large volume of healthcare data can be computed efficiently by AI algorithms, in order to assist clinical practice. Intelligent systems can be equipped with learning and self-correcting skills to improve their accuracy, and support physicians in reducing diagnostic and therapeutic errors. Moreover, AI systems can be used so as to extract information from a large patient population and assist in making real-time inferences for a health risk alert and outcome prediction[14].

AI medical devices can be classified into two categories: *a)* ML techniques that analyse healthcare data (e.g., imaging, genetic and EP data) in order to cluster patients' characteristics or infer the probability of the outcome of certain diseases; *b)* natural

---

[10] Within the **genus** «therapeutics», three *species* can be distinguished: *a)* medical therapy; *b)* surgical therapy; *c)* mixed therapy (e.g., some neoplasia needs to be treated by means of: first, an oncological intervention, so as to reduce the **critical mass**; secondly, a surgical intervention to remove it; and finally, another oncological intervention in order to prevent the risk of metastasis).

[11] However, recent studies have compared the level of diagnostic efficiency of humans and AI systems. The outcome was the humans' **victory**. The spread between the samples is directly proportional to the rarity of diseases: the more a disease is atypical, the more humans take advantage of their personal skills, since the clinical question cannot be simply solved by means of statistics. Instead, when a disease is common, AI systems can use their (potentially boundless) skills in data processing. See H.L. Semigran, D.M. Levine, S. Nundy, A. Mehrotra, *Comparison of Physician and Computer Diagnostic Accuracy*, in *JAMA Internal Medicine*, 176, 12, 2016, pp. 1860-1861.

[12] S. Saria, A. Butte, A. Sheikh, *Better medicine through machine learning: What's real, and what's artificial?*, in *PloS Med*, 15, 12, 2018, p. 1.

[13] E.H. Shortliffe, M.J. Sepúlveda, *Ibidem*.

[14] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, *Artificial intelligence in healthcare: past, present and future*, in *Stroke and Vascular Neurology* 2, 4, 2017, p. 230.

language processing (NLP) methods that process unstructured data (e.g., clinical notes and medical journals) in order to develop structured medical data. In particular, DL is very performing in the interpretation of data in the form of images, by virtue of the complexity of factors that it can take into account.

Medical literature has widely discussed the advantages of AI, mainly regarding three disease types: cardiovascular disease, cancer and nervous system disease. These are the leading causes of death. Thus, early diagnoses are essential in order to avert the decline of patients' health status. This can be achieved through improving the analysis procedures on healthcare data by means of AI techniques.

Having said this, several challenges still need to be faced. In the first place, current regulations lack standards to assess the AI systems' safety and efficacy. Furthermore, scientists deal with problems of data-exchange. From this point of view, one should consider that intelligent systems need to be constantly trained by clinical data: the first training can be based on historical datasets; however, the following steps require a continuation of the information. This is a crucial issue for the development and improvement of the system[15].

4. Criminal Law Questions Raised by the Use of AI in Healthcare.

Many questions in the field of criminal law arise from the development of software as a medical device (SaMD). In brief, we need to clarify:

a) whether AI systems must be considered as **agents**, in as much as they can be **legal persons**, or mere **instruments**, through which humans might commit crimes;
b) how AI crimes (AIC) – i.e. crimes involving an AI system – can be performed in the field of medical devices, namely whether they are crimes typically based on specific conduct or requiring a specific event to occur[16]; also, in the latter case, how we can solve the questions concerning the etiological link between the agent's conduct and the event, and the varieties of fault; and
c) finally, whether a new fairer model may be theorised.

5. AI Entities as Legal Persons.

Gabriel Hallevy theorised possible forms of AI systems' criminal liability based on the attribution to them of a legal personality. He postulated three models:

i. the **perpetration-through-another** model, where the AI system is considered as an **innocent agent**, a mere **instrument** used by the actual perpetrator (**principal**), i.e. the programmer or the user;

[15] F Jiang *et al.*, *Ivi*, p. 241.
[16] T.C. King, N. Aggarwal, M. Taddeo, L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *Sci. Eng. Ethics*, 26, 2020, pp. 90-91.

ii. the **natural-probable-consequence** model, where programmers/users may be held criminally liable for a crime committed via AI and occurring as a natural and probable consequence of their intentional or negligent behaviour;
iii. the **direct liability** model, where it is assumed that AI is endowed with *mens rea*, and therefore compatible with a burden of liability for conduct materially executed by itself.

In the last model, Hallevy hypothesises a possible application of punishment constructed according to a principle of equivalence between machine and man. It would be a matter of eliminating the software, intending to neutralise the system, or deactivating it for a pre-established length of time in order to encourage its re-education. Yet, Hallevy's third model cannot be accepted. It is based on vitiated arguments.

First, AI systems are not actually **intelligent**. In law, speculation without scientific evidence cannot constitute a valid reference. As Luciano Floridi, a professor at the Internet Oxford Institute, argues, the best definition of AI is still the one provided by John McCarthy in 1955: the AI problem «is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving»[17]. Thus, we can call such a behaviour intelligent in as much as a human behaves in that way, but it does not mean that the machine is intelligent[18].

Hallevy argues that the objections against the criminal liability of AI entities (above all concerning the requirement of *mens rea*) are based on arguments that are similar to the ones relating to the liability of corporations. Then, there would be «no substantial legal difference between the idea of criminal liability imposed on corporations and on AI entities». Yet, there is a substantial difference: «under the […] 'superior agent' rule, corporate criminal liability […] is limited to situations in which the conduct is performed or participated in by the board of directors or a high managerial agent»[19]. Instead, Hallevy's third model «does not assume any dependence of the AI entity on a specific programmer or user»[20].

Finally, criminal responsibility is based on the two crucial concepts of **wrongdoing** and **attribution**[21], which presuppose two requisites:

---

[17] J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence; August 31, 1955*, in *AI Magazine* 27, 4, 2006, p. 12.
[18] L Floridi, *Digital's Cleaving Power and Its Consequences*, in *Philos. Technol.*, 30, 2017, pp. 123 ff.
The author argues that if one affirmed that the machine is **intelligent**, it would be «a fallacy that smacks of superstition (compare: the river reaches the lake by following the best possible path, removing obstacles in its way; if this had been done by someone, we would have considered that behaviour intelligent; so the river's behaviour is intelligent)».
[19] W.R. LaFave, A.W. Scott Jr., *Substantive Criminal Law*, Volume 1, West Publishing Co., 1986, p. 360.
[20] G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4, 2, Article 1, 2010, p. 186.
[21] An act is **wrongful** when it satisfies the definition of an offence and is unjustified. The definition of an offence is composed by a set of (objective and subjective) elements that constitutes the so-called **incriminating** *case* against the charged person. It is provided for in a *basic* norm, i.e., a norm, directed to the citizens, which **prohibits** particular acts or **requires** particular acts to be performed. So, they impose **duties of compliance** on individuals. Therefore, the concept of **wrongful conduct** is formal, being defined by the incompatibility of the act with respect to the norms of the legal system. Instead, the concept of **wrongdoing** is substantial: to say

a) a **human** (and **voluntarily taken**) act or omission[22] (also known as *actus reus*)[23];
b) the *mens rea* (varieties of fault), i.e., the act/omission needs to be *covered* by a guilt mind.

6. AI Entities as Instruments: Responsibility of the *Producers* and the *Users*.

We have clarified that AI systems are not **subjects**. Instead, they have to be considered as **instruments**, through which the actor can commit a crime. In the field of AI medical devices, we can distinguish two types of offences: the ones committed by the **producers** and the ones committed by the **users**.

Producers can be found liable for having committed crimes either within the pattern of **manifest criminality**, or within the pattern of **harmful consequences**. In the first case, «the significance of the criminal act is that it manifests criminality and unnerves the community's sense of security»[24]. In other words, what is merely punished is that the actor puts in **danger** a legal interest (e.g., life, or public health). These offences are forms of **direct** liability that can consist of either **active** or **omissive** conduct (the breach of a specific statutory duty to act), and are characterised by considerable anticipation of the threshold to trigger criminal law protection. In this respect, we may generally distinguish:

a) the responsibility for the **type** of production, when the perpetrator develops certain types of AI devices **absolutely prohibited** by the law, since they involve a high risk of causing harmful events, that cannot be (substantially) reduced through the respect of some precautionary measures; and
b) the responsibility for the **modality** of production, that relates to areas of production (**allowed** by the law) characterised by a lower risk-coefficient, since the available nomological skills permit to set precautionary measures in order to prevent/reduce the risk of harmful events[25]. Thus, producers can be held accountable for having produced AI devices without complying with the aforementioned precautionary rules.

---

that the perpetrator is a **wrongdoer** (or that he/she engaged in wrongdoing) is to pass judgment on the intrinsic quality of his/her deeds. The concept of **attribution** can be referred to in two dimensions:
– **objective** attribution is the term used in Germany to qualify the general process of holding individuals accountable for the occurrence of harm or the acts of other persons; and
– **subjective** attribution, instead, refers to the distinct question of the **criteria for holding persons accountable** for their deeds.
With particular reference to **subjective** attribution, it is based on norms other than *basic* ones, directed (not to the citizens but) to the judge. These rules do not generate exceptions to the basic norms, they only **excuse** their violation.
In this regard, see *amplius* the fundamental work by G.P. Fletcher, *Rethinking Criminal Law*, Oxford University Press, 2000, pp. 454 ff..
[22] G.P. Fletcher, *Ivi*, p. 475.
[23] T.C. King *et al.*, *Ibidem*.
[24] G.P. Fletcher, *Ivi*, p. 420.
[25] F. Bricola, *Responsabilità penale per il tipo e per il modo di produzione*, in *Scritti di diritto penale*, 1, II, ed. S. Canestrari, A. Melchionda, Giuffrè, 1997; C Piergallini, *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*, Giuffrè, 2004, pp. 40 ff.

Furthermore, they can be found guilty of a crime within the pattern of **harmful consequences**, e.g., manslaughter, in as much as, within the productive cycle of an AI device, they have not respected the precautionary rules mentioned above, and, because of this violation, they have caused the **actual occurrence of a harm** to the legal interest protected by the law (individuals' life), such as a man/woman's death.

With regard to the **users**, namely physicians, they can commit crimes within the aforementioned patterns too. As for offences within the pattern of **manifest criminality**, in the field of SaMD, these forms of responsibility can consist in violating the ban of merely using devices with certain characteristics, i.e., involving a risk-level that is considered excessive and unacceptable by the law. As for crimes within the pattern of **harmful consequences**, physicians can be found liable for having committed manslaughter due to **medical malpractice**. These offences could fulfil forms of:

a) **direct** liability, when the physician commits **active** deeds, i.e., he/she uses a SaMD, and, as a **direct** consequence, the patient dies; and
b) **derivative** liability, when he/she is responsible for a so-called **commission by omission**.

Indeed, the paradigm of **commission by omission** is based on the criteria of the so-called **duty to avert the harmful event**. The duty to act – a **legal** duty, not merely a **moral** duty – can arise according to:

i. a statute («other than the criminal statute whose violation is in question»)[26];
ii. a personal relationship, e.g., between the physician and the patient;
iii. a contract;
iv. the voluntary assumption of care; and
v. the creation of a risk by the defendant.

In other words, the physician can be held accountable for the event that is to be averted only if – in the light of a special **protection-link** –, in a particular situation, he/she can be considered as the **defender** (the **guarantor**) of a certain legal interest, namely a patient's life[27]. For instance: a physician used a SaMD; he/she had the **duty to control** it, so as to prevent (**as much as possible**) the occurrence of harmful outcomes; nevertheless, he/she failed in doing it (either because he/she **absolutely** did not control the **causal progress** activated by the device's use, or because he/she **erroneously** controlled it, namely not complying with the applicable guidelines). According to the prevailing view, the perpetrator must know the facts indicating a duty to act, and the jurisprudence retains that the imposition of strict liability in omission cases is inappropriate[28]. Certainly, «sometimes there may be a duty to take care to know the facts, as well as a duty to go into action when the facts are known»[29]. Also, the actor must be physically capable of performing the actions necessary to avert the harmful event[30]. In the pattern of **harmful consequences**,

[26] W.R. LaFave, A.W. Scott Jr., *Ivi*, 286.
[27] This element is common to several legal systems. In this regard, one could mention the Italian notion of "*posizione di garanzia*", or the German concept of "*Garantenstellung*".
[28] *Harding v. Price*, 1948, 1 K.B. 695.
[29] W.R. LaFave, A.W. Scott Jr., *Ivi*, pp. 290-291.
[30] W.R. LaFave, A.W. Scott Jr., *Ivi*, p. 291.

when an event occurs due to the use of AI systems, a problem might arise with reference to causation. As things stand, we are not able to identify the etiological links within the neural nets in terms of certainty. In other words, one can affirm that an AI system causes a certain event *y* (e.g., a diagnosis), from a certain data-set *x* (e.g., some tests on the patient). However, the reconstruction of every single intermediate step through which the system reaches that result (especially in cases of DL) is currently impossible.

Nonetheless, by means of the counterfactual analysis, judges only need to establish whether a certain conduct is a contingently essential condition of the event. They never achieve a result that is deductively sure. Indeed, their reasonings are almost always probabilistic (i.e., rationally plausible), due to the lack of explicative preconditions and to the use of statistical laws in the explanation of naturalistic events[31]. In this sense, as a matter of common knowledge, the actor's conduct, namely the use of a SaMD, must be the *condicio sine qua non* of the harmful event[32], or that but for the conduct, the event would not have occurred[33].

In commission-by-omission cases, the judge must:

i. reconstruct the omitted action[34], i.e., the (correct) supervision of the device that would have avoided the deterioration of a patient's health condition;
ii. reconstruct the causal process, as it actually occurred, i.e., how and why the harmful event occurred; and
iii. clarify – in the light of an established scientific law – whether the actor's (hypothetical) conduct would have affected the causal process averting the harmful event.

A point must be stressed: the standard of proof expressed by the formula beyond any reasonable doubt (*alias* "BARD rule") arises from the consideration that «it is far worse to convict an innocent man than to let a guilty man go free»[35]. Therefore, in this field, an established scientific covering-law is essential to prove the etiological process and then the actor's liability. Otherwise, the accused physician must be declared not guilty.

---

[31] F. Stella, *La nozione penalmente rilevante di causa: la condizione necessaria*, in *Rivista Italiana di Diritto e Procedura Penale*, 4, 1988, p. 1217.
[32] For the purpose of criminal law, one can affirm that there is an etiological link between an antecedent (e.g., a – wrong – diagnosis obtained by means of an AI system) and an event (e.g., the patient's death) when two elements subsist, namely:
i. an established (universal or statistical) covering-law, in the light of which the causal process in question can be explained; and
ii. the agent's conduct (i.e., the use of an AI system) is an essential condition in all the feasible (or probable) explanations.
[33] The American Law Institute, *Model Penal Code*, § 2.03, 1, May 24, 1962.
[34] G. Grasso, *Il reato omissivo improprio: La struttura obiettiva della fattispecie*, Giuffrè, 1983, pp. 370-371.
[35] *In re Winship*, 397 U.S. 358, 1970; *Barnes v. United States*, 412 U.S. 873, 1973; *Carella v. California*, 491 U.S. 263; *Albright v. Oliver et al.*, 510 U.S. 266, 1994; *Victor v. Nebraska*, 511 U.S. 1, 1994; *United States v. Gaudin*, 515 U.S. 506; *Spencer v. Kemna*, 523 U.S. 1, 1998. As for the jurisprudence of the Italian Supreme Court see above all: Cassazione Penale, Sezioni Unite, no. 30328 *Franzese*, 2002.

In order for the crime to be fulfilled, the perpetrator must meet the so-called *mens rea* requirements. In this regard, the offences can be basically committed in the light of the following varieties of fault[36]:

a) intention;
b) knowledge;
c) recklessness;
d) negligence.

Among these varieties of fault, **recklessness** and **negligence** are the most questionable, relating to the notion of **risk** (that must be **substantial** and **unjustifiable**)[37]. The actor – respectively – consciously has disregarded it, or should have been aware of it[38]. The threshold of substantiality and unjustifiability of the risk is due to a *gross* deviation from the standard of conduct/care[39] that a **law-abiding/reasonable person**[40] would respect in the actor's situation[41].

---

[36] One might observe that the various legal systems adopt different solutions concerning the varieties of fault. As a matter of fact, this consideration is only partially true. Instead, the comparative analysis demonstrates that, regardless of terminological differences, substantial similarities are often more than dissimilarities. Sure, solutions are rarely the same. Nevertheless, modern jurists' task – living in a globalised world – is to **build bridges not walls**, to find points of contact between different legal traditions, especially in areas – such as AI – where criminality (and its prevention and punishment) no longer concerns one State solely, but requires stronger mechanisms of judicial cooperation.
From this point of view, one might refer to the rules and principles set in the Rome Statute of the International Criminal Court (where applicable). Sure, the ICC exercises its jurisdiction over crimes that are other than the ones we are dealing with, i.e., artificial intelligence crimes ("AIC") in the field of healthcare. However, some definitions – especially the ones concerning the so-called **general part** of criminal law – may represent valid references. As a matter of fact, the Statute's drafters (most of which were comparative law experts) made the effort to formulate definitions that could have been a sort of synthesis between the various legal traditions.
In this regard, for example, one might mention the definition of *mens rea*, based on the elements of **intent** and **knowledge**, pursuant to article 30: «[…] a person has intent where: (a) In relation to conduct, that person means to engage in the conduct; In relation to a consequence, that person means to cause that consequence or is aware that it will occur in the ordinary course of events. […] 'knowledge' means awareness that a circumstance exists or a consequence will occur in the ordinary course of events […]».
[37] The American Law Institute, *Model Penal Code*, § 2.02(2)(d): «A person acts negligently with respect to a material element of an offense when he should be aware of a substantial and unjustifiable risk that the material element exists or will result from his conduct. The risk must be of such a nature and degree that the actor's failure to perceive it, considering the nature and purpose of his conduct and the circumstances known to him, involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation».
[38] In this respect, we should distinguish three elements: *a)* the **recognisability** ("*Erkennbarkeit*" in Germany, "*riconoscibilità*" in Italy) of the unreasonable/unjustifiable risk; *b)* its **foreseeability** ("*Vorhersehbarkeit*" in Germany, "*prevedibilità*" in Italy); and *c)* its **preventability** ("*Vermeidbarkeit*" in Germany, "*evitabilità*" in Italy).
[39] The notion of «gross deviation from the standard of care» expresses what is commonly understood as the **something extra** that distinguishes the **criminal** (or **gross**) negligence and the **ordinary** negligence, which is relevant in the area of tort law solely.
Analogous (albeit not identical) concepts can be found in the German and Italian legal systems: i.e., the notions of – respectively – "*Leichtfertigkeit*" and "*colpa grave*".
[40] W.R. LaFave, A.W. Scott Jr., *Ivi*, p. 328: «Thus negligence is framed in terms of an objective (sometimes called 'external') standard, rather than in terms of a subjective standard».
[41] In this regard, see also W.R. LaFave, A.W. Scott Jr., *Ivi*, pp. 327-328: «"Unreasonable risk" is an expression which takes into account the fact that we all create some risk to others in our everyday affairs without subjecting ourselves to liability for negligence […]. The test for reasonableness in creating risk is thus said to be determined by weighing the magnitude of the risk of harm against the utility of the actor's conduct. Under such a test, even a slight risk may be unreasonable […]. Aside from the utility of the actor's conduct, another

With regards to the aforementioned standard, some rules of care are crucial. As for **producers**, norms on the rigorous clinical validation of AI medical devices can be cited[42]. Instead, the standard of care for **users**, i.e., physicians, guidelines and best practices on healthcare are the main reference. However, in both cases, the judge must investigate whether, in the specific situation, the rule of care that the actor violated was aimed at preventing harmful events of the type of the one that occurred[43].

7. A New Paradigm.

The abovementioned liability paradigm (still in effect) does not allow to realise the simultaneous **protection of the innocent and the victims**[44]. Another paradigm can be theorised in a *de lege ferenda* perspective.

In the comparative law-panorama, the U.S. model of assessment of technological risks can be a good reference. It is based on the central role of public agencies (e.g., the Food and Drugs Administration) and on five elements[45]:

i. The agency provides for specific **precautionary rules** that, for example, the companies involved in the production/use of medical devices must respect. These rules set risk-assessment **standards**, i.e. the adequacy threshold of a risk;

---

variable factor involved in the question of whether a particular risk is unreasonable is the extent of the actor's knowledge of the facts bearing upon the risk […]. Still another variable concerns the nature and the extent of the harm which may be caused by the defendant's conduct, and the number of persons who may be harmed».

[42] The International Medical Device Regulators Forum (IMDRF) – of which the WHO is an official observer within the management committee –, in particular the SaMD Working Group, has developed a series of documents in order to provide harmonized principles for individual jurisdictions to adopt based on their own regulatory framework.
The US Food & Drugs Administration (FDA) has adopted these principles. In this respect, see IMDRF SaMD Working Group, *Software as a Medical Device (SaMD): Key Definitions*, December 9, 2013; Id., *Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations*, September 18, 2014; Id., *Software as a Medical Device (SaMD): Application of Quality Management System*, October 2, 2015; Id., *Software as a Medical Device (SaMD): Clinical Evaluation*, June 22, 2017.
In Europe, on the 5th April 2017 the European Parliament and the Council of the European Union has adopted the Regulation (EU) 2017/745 **on medical devices**. This «lays down rules concerning the placing on the market, making available on the market or putting into service of medical devices, for human use and accessories for such devices in the Union. This Regulation also applies to clinical investigations concerning such medical devices and accessories conducted in the Union» (article 1.1). The abovementioned guidance developed for medical devices at the international level have been taken into consideration by the EU too, so as «to promote the global convergence of regulations which contributes to a high level of safety protection worldwide, and to facilitate trade» (recital 5). In this way, the results of clinical investigations conducted in the EU would be accepted as documentation outside the EU, and those conducted outside the EU in accordance with international guidelines would be accepted within the EU. In this regard, the rules should be in line with the most recent version of the World Medical Association (WMA) Declaration of Helsinki *on ethical principles for medical research involving human subjects* (recital 64).

[43] In Germany and Italy this concept is understood as, respectively, "*Risikozusammenhang*" and "*nesso di rischio*". These expressions' translation might be the **risk-link** between the violation of the duty of care (by the actor) and the harmful event, whose nature is normative, not material.

[44] F. Stella, *Giustizia e modernità: La protezione dell'innocente e la tutela delle vittime*, Giuffrè, 2003.

[45] S. Jasanoff, *Science at the Bar: Law, Science, and Technology in America*, Harvard University Press, 1995, pp. 69 ff.; F. Centonze, *La normalità dei disastri tecnologici: Il problema del congedo dal diritto penale*, Giuffrè, 2004, pp. 400 ff.

ii. The adoption of these precautionary rules is **democratically legitimated** (i.e., the Congress approves them);
iii. The courts review the rationality of the agencies' decisions, i.e., courts of appeals may invalidate them in case they are **arbitrary** or **capricious**, or not supported by **substantial evidence**[46];
iv. A rigid and preventive **enforcement** system, i.e., agencies may make **inspections** of the companies and issue **injunctions** to them;
v. An **education and compliance assistance** system, i.e., the companies can either ask for the agency's **consultation assistance** in arranging an adequate set of cares able to prevent harmful events or accept a **Voluntary Protection Program** in the light of which the agency itself implements a permanent **on-site analysis** on the internal safety-system.

From a broader perspective, international harmonisation/cooperation is desirable in this field. The World Health Organization (WHO) might set general standards, that national – or hopefully supranational agencies, such as the European Medicines Agency (EMA) – should (not might) subsequently implement into more specific guidelines for producers and users.

As for sanctions, a multi-level system can be theorised. Above all, we should distinguish sanctions for the liability of organisations and sanctions for the liability of individuals.

As for the former, two levels can be hypothesised:
i. Administrative sanctions (imposed by an agency) for minor violations of the precautionary rules concerning the production/use of a SaMD;
ii. Criminal sanctions (imposed by the courts) for serious violations of the aforementioned precautionary rules and, eventually, for causing harmful events to the patients because of such violation.

As for the individuals, the view is necessarily different: in a field characterised by **scientific uncertainty**, they might be unconscious that, through their deeds, they create a substantial and unjustifiable risk. Thus, their liability should be limited to conducts covered by the psychological coefficients of **intent** and, at most, of **recklessness**. **Negligent** deeds should remain, in this field, immune from penal sanctions[47].

Finally, two considerations need to be pointed out. First, regardless of whether the accused is an organisation or an individual, the evidentiary and judgment rule in criminal trials is the one expressed by the formula **beyond any reasonable doubt**. Then, in cases where the prosecutors cannot present sufficient evidence so as to satisfy the aforementioned standard (either because they do not find it, or because – especially with

[46] *US Administrative Procedure Act*, 1946, Section 10.
[47] T.C. King *et al.*, *Ivi*, 95: «Concerning the knowledge threshold, in some cases the *mens rea* could actually be missing entirely. The potential absence of a knowledge-based *mens rea* is due to the fact that, even if it is understood that an AA [artificial agent] can perform the *actus reus* autonomously, the complexity of the AA's programming makes it possible that the designer, developer, or deployer (i.e., a human agent) will neither know nor predict the AA's criminal act or omission. The implication is that the complexity of AI provides a great incentive for human agents to avoid finding out what precisely the ML [machine learning] system is doing, since the less the human agents know, the more they will be able to deny liability for both these reasons».

reference to individuals' liability – it does not exist), victims can follow the path of tort law, in order to obtain fair compensation for the damages they suffered. In this regard, the view of those who theorise definitions of criminal liability without fault requirement cannot be accepted, and *mens rea* is so crucial to the State's entitlement to sanction individuals depriving them of their freedom that we cannot merely abandon it only because of difficulty in proving it.

8. Conclusions.

To sum up, we have seen that AI applications in healthcare entail both benefits and risks, which need to be balanced. The search for an equilibrium between them involves considerations concerning not only the field of medicine, but also ethics and law.

In particular, the protection of human rights deserves attention. Among them, in the field of criminal law, one should stress the interconnection that exists between, on the one hand, the principles of legality (*nullum crimen sine lege*) and of culpability (*nullum crimen sine culpa*), and, on the other hand, their processual *pendant*, namely the presumption of innocence and the BARD rule (a rule relating both to the admission and evaluation of evidence, and to the verdict on the defendant's guilty). From the point of view of the accused, the observance of these principles is crucial so as to guarantee the fundamental right to a fair trial[48]. Nevertheless, patients-victims deserve justice in case of damages caused by mistakes in applying AI techniques to their clinical situation. We have seen how this simultaneous protection of the innocent and of the victims can be reached. In the first place, one should not confuse the liability of producers, and the one of physicians. The former can be found responsible for having committed offences either within the pattern of manifest criminality (i.e., they violate the absolute ban of producing certain types of devices, or they develop devices without complying with the applicable precautionary rules) or within the pattern of harmful consequences (in so far as the violation of the precautionary is etiologically linked to a man/woman's death). Instead, the case of physicians' liability is essentially a matter of medical malpractice. Then, from a different point of view, one should distinguish the cases of liability of individuals and legal persons, given that several legislations can differently apply to each of them. Finally, in a *de lege ferenda* perspective, we have seen that a new paradigm can be theorised. It is inspired to the US model of risk assessment, and involves the integration of criminal law, civil law, and administrative law measures. I think that this model could guarantee an equilibrium between the fundamental rights of the accused (a fair trial) and of the victims (a compensation for damages they have suffered).

[48] G. Grasso, *La protezione dei diritti fondamentali nella Costituzione per l'Europa e il diritto penale: spunti di riflessione critica*, in *Lezioni di diritto penale europeo*, ed. G. Grasso, R. Sicurella, Giuffrè, 2007, pp. 656 ff.

Bibliography

Books

M.A. Boden, *Artificial Intelligence. A Very Short Introduction*, Oxford University Press, 2018, (Italian trans. *L'intelligenza artificiale*, Il Mulino, 2019).

F. Bricola, *Responsabilità penale per il tipo e per il modo di produzione*, in *Scritti di diritto penale*, 1, II, ed. S, Canestrari, A. Melchionda, Giuffrè, 1997.

F. Centonze, *La normalità dei disastri tecnologici: Il problema del congedo dal diritto penale*, Giuffrè, 2004.

G.P. Fletcher, *Rethinking Criminal Law*, Oxford University Press, 2000.

G. Grasso, *Il reato omissivo improprio: La struttura obiettiva della fattispecie*, Giuffrè, 1983.

G. Grasso, *La protezione dei diritti fondamentali nella Costituzione per l'Europa e il diritto penale: spunti di riflessione critica*, in *Lezioni di diritto penale europeo*, ed. G. Grasso, R. Sicurella, Giuffrè, 2007.

S. Jasanoff, *Science at the Bar: Law, Science, and Technology in America*, Harvard University Press, 1995.

W.R. LaFave, A.W. Scott Jr., *Substantive Criminal Law*, Volume 1, West Publishing Co., 1986.

C Piergallini, *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*, Giuffrè, 2004.

F. Stella, *Giustizia e modernità: La protezione dell'innocente e la tutela delle vittime*, Giuffrè, 2003.

Journal Articles

C. Casonato, *Potenzialità e sfide dell'intelligenza artificiale*, in *BioLaw Journal*, 1, 2019, pp. 177 ff.

L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, *AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and machines*, 28, 4, 2018, pp. 689 ff.

L Floridi, *Digital's Cleaving Power and Its Consequences*, in *Philos. Technol.*, 30, 2017, pp. 123 ff.

G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4, 2, Article 1, 2010, p. 186.

F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, *Artificial intelligence in healthcare: past, present and future*, in *Stroke and Vascular Neurology* 2, 4, 2017, p. 230.

T.C. King, N. Aggarwal, M. Taddeo, L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *Sci. Eng. Ethics*, 26, 2020, pp. 90-91.

R.S. Ledley, L.B. Lusted, *Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason*, in *Science*, 130, 3366, July 3, 1959.

J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence; August 31, 1955*, in *AI Magazine* 27, 4, 2006, p. 12.

Y.-S. Park, S. Lek, *Artificial Neural Networks: Multilayer Perceptron for Ecological Modelling*, in *Developments in Environmental Modelling*, 28, 2016, p. 124.

M.J. Rigby, *Ethical Dimensions of Using Artificial Intelligence in Health Care*, in *AMA J Ethics*, 21, 2, 2019, pp. 121 ff.

S. Saria, A. Butte, A. Sheikh, *Better medicine through machine learning: What's real, and what's artificial?*, in *PloS Med*, 15, 12, 2018, p. 1.

H.L. Semigran, D.M. Levine, S. Nundy, A. Mehrotra, *Comparison of Physician and Computer Diagnostic Accuracy*, in *JAMA Internal Medicine*, 176, 12, 2016, pp. 1860-1861.

E.H. Shortliffe, M.J. Sepúlveda, *Clinical Decision Support in the Era of Artificial Intelligence*, in *JAMA*, 320, 21, December 4, 2018, p. 2199.

F. Stella, *La nozione penalmente rilevante di causa: la condizione necessaria*, in *Rivista Italiana di Diritto e Procedura Penale*, 4, 1988, pp. 1217 ff.

Institutional Documents (Statutes, Jurisprudence, Other Official Documents)

Albright v. Oliver et al., 510 U.S. 266, 1994.

Barnes v. United States, 412 U.S. 873, 1973.

Carella v. California, 491 U.S. 263.

Cassazione Penale, Sezioni Unite. no. 30328 *Franzese*, 2002.

E.U. Commission, *Artificial intelligence for Europe*, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2, April 25, 2018.

EU High-Level Expert Group on AI, *Ethic Guidelines for Trustworthy AI*, 4, April 8, 2019.

Harding v. Price, 1948, 1 K.B. 695.

IMDRF SaMD Working Group, *Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations*, September 18, 2014.

IMDRF SaMD Working Group, *Software as a Medical Device (SaMD): Application of Quality Management System*, October 2, 2015.

IMDRF SaMD Working Group, *Software as a Medical Device (SaMD): Clinical Evaluation*, June 22, 2017.

In re Winship, 397 U.S. 358, 1970.

Spencer v. Kemna, 523 U.S. 1, 1998.

The American Law Institute, *Model Penal Code*, § 2.03(1), May 24, 1962.

The European Parliament, and the Council of the European Union, *Regulation (EU) 2017/745 on medical devices*, April 5, 2017.

U.S. Administrative Procedure Act, 1946.

United States v. Gaudin, 515 U.S. 506.

Victor v. Nebraska, 511 U.S. 1, 1994.