AI Systems Involved in Harmful Events: Liable Persons or Mere Instruments? An Interdisciplinary and Comparative Analysis

Federico Carmelo La Vattiata*

ABSTRACT: The article investigates the nature of AIs under criminal law, *i.e.*, whether they are legal persons or mere tools. The study applies a double methodology. Firstly, from a comparative perspective, it analyses the US and the Italian legal systems, as they represent the two main legal traditions in the Western World, namely, common law, and civil law. Secondly, it applies the interdisciplinary research method, by reference to non-legal disciplines. The article criticizes the doctrine maintaining that AIs may be considered as legal persons. Then, it aims to demonstrate the opposite thesis, according to which AIs are mere tools.

KEYWORDS: AI; Criminal Law; Legal Persons; Tools

SUMMARY: 1. Introduction – 2. Methodology – 3. Literature Review – 4. Facts' Overview and Discussion – 4.1. Trustworthy AI – 4.2. Automation versus Autonomy – 4.3. Thesis: AIs are legal subjects – 4.4. Antithesis: AIs are mere objects – 4.4.1. Is AI actually 'intelligent'? – 4.4.2. The substantial legal difference between corporations and AI systems – 4.4.3. *Machina delinquere et puniri non potest* – 5. Recommendations and Conclusive Remarks.

1. Intoduction

Artificial intelligence (AI) systems[1] are part of our lives. For instance, autonomous vehicles (AV) already circulate on the roads.[2] Also, physicians daily apply AI-based medical devices.[3] In either case, AI systems can be involved, respectively, in road accidents and in patients' deaths and injuries. Hence, new regulatory challenges arise.[4]

* *PhD in Comparative and European Legal Studies (specialisation: Criminal Law and Procedure and Philosophy of Law), University of Trento. Mail: federicolavattiata@gmail.com.* *The article was subject to a double-bind peer review process.*

[1] See S.J. RUSSELL, P. NORVIG, *Artificial intelligence: a modern approach*, Hoboken, 2021.

[2] X. DI, R. SHI, *A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning*, in *Transportation Research Part C: Emerging Technologies*, 125, 2021, 103008; M. MARTÍNEZ-DÍAZ, F. SORIGUERA, *Autonomous vehicles: theoretical and practical challenges*, in *Transportation Research Procedia*, 33, 2018, 275-282.

[3] D.B. KRAMER, S. XU, A.S. KESSELHEIM, *Regulation of Medical Devices in the United States and European Union*, in *A.L.* CAPLAN, B. PARENT (eds.), *The Ethical Challenges of Emerging Medical Technologies*, Abingdon/New York, 2017, 41-48; G. TRIFIRÒ, S. CRISAFULLI, G. PUGLISI, G. RACAGNI, L. PANI, *Terapie digitali come farmaci?*, in *Tendenze nuove*, 1, 2021, 147-158; G. RECCHIA, D.M. CAPUANO, N. MISTRI, R. VERNA, *Digital Therapeutics-What they are, what they will be*, in *Acta Scientific Medical Sciences*, 4, 2020, 134-142.

[4] Đ. PETROVIĆ, R. MIJAILOVIĆ, D. PEŠIĆ, *Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers*, in *Transportation Research Procedia*, 45, 2020, 161-168; V.V. DIXIT, S. CHAND, D.J. NAIR, *Autonomous Vehicles: Disengagements, Accidents and Reaction Times*, in *PLoS ONE*, 11, 2016,

This survey addresses the following question: when a person dies or suffers from injuries because of an AI device, who should be liable for the occurrence of such event? Can the AI system itself be considered the perpetrator of murder, manslaughter, battery, or assault in the US,[5] and of *omicidio* or *lesioni personali*[6] in Italy?[7] Or should it be considered as a mere tool, since the actual offender is one of the various economic actors involved in its lifecycle (who, for example, failed to comply with the duty to control it)?

2. Methodology

It has previously been written about some of the questions addressed in this Article.[8] They will be scrutinized further herein.

The interdisciplinary research method will be adopted. Especially when complex questions – like the regulation of AI – are at stake, jurists should abandon the logic of autopoiesis, that is, the characteristic of social subsystems (*e.g.*, law, economics, science, and so on) with a self-contained and a self-moving nature.[9] Rather, they should incorporate insights from non-legal disciplines into their studies.[10] Problems often transcend the boundaries of a particular discipline.[11] Problems – not disciplines – need to be studied so that solutions can be found.[12]

---

e0168054; W. Nicholson Price II, *Regulating Black-Box Medicine*, in *Michigan Law Review*, 116, 3, 2017, 421-474; W. Nicholson Price II, *Medical Malpractice and Black-Box Medicine*, in I.G. Cohen, H.F. Lynch, E. Vayena, U. Gasser (eds.), *Big Data, Health Law, and Bioethics*, Cambridge (UK)/New York, 2018, 295-306; W. Nicholson Price II, *Artificial Intelligence in the Medical System: Four Roles for Potential Transformation*, in *Yale Journal of Law & Technology*, 21, 2019, 122-132; M.U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, in *Harvard Journal of Law & Technology*, 29, 2, 2016, 353-400.

[5] See American Law Institute, *U.S. Model Penal Code*, § 210.0, § 210.1, § 210.2, § 210.3, § 210.4, § 211.0, § 211.1 (Proposed Official Draft 1962).

[6] See Article 575 *et seq*. of the Italian Criminal Code.

[7] W.R. LaFave, A.W. Scott, *Substantive criminal law*, 2, St. Paul (Minn., US), 1986, 179-323; G.P. Fletcher, *Rethinking criminal law*, Oxford/New York, 2000, 325-340; J.D. Ohlin, *Criminal law: doctrine, application, and practice*, New York, 2016, 213-324; J. Herring, *Criminal Law: Text, Cases, and Materials*, New York, 2020, 237-419; G. Fiandaca, E. Musco, *Diritto penale. 2,1: Parte speciale I delitti contro la persona*, Bologna, 2020, 94-172.

[8] F.C. La Vattiata, *Artificial Intelligence in Healthcare: Risk Assessment and Criminal Law*, in I. Beridze, S. van de Meulengraaf, O. McCarthy, A. Rodriguez Tamayo (eds.), *UNICRI Special Collection on Artificial Intelligence*, Torino, 2020, 48-58.

[9] N. Luhmann, *The Autopoiesis of Social Systems*, in R.F. Geyer, J. van der Zouwen (eds.), *Sociocybernetic paradoxes: observation, control, and evolution of self-steering systems*, London/Beverly Hills, 1986, 172-192; A. Lourenco, *Autopoietic social systems theory: the co-evolution of law and the economy*, in *Australasian Journal of Legal Philosophy*, 35, 2010, 35-54.

[10] W. Schrama, *How to carry out interdisciplinary legal research. Some experiences with an interdisciplinary research method*, in Utrecht Law Review, 7, 2011, 147-162; D.W. Vick, *Interdisciplinarity and the Discipline of Law*, 31, 2, 2004, 163-193; K.M. Sullivan, *Foreword: Interdisciplinarity,* in *Michigan Law Review*, 100, 2, 2002, 1217-1226.

[11] K.R. Popper, *Conjectures and refutations: the growth of scientific knowledge,* London/New York, 2002; F. Stella, *Giustizia e modernità: la protezione dell'innocente e la tutela delle vittime*, Milano, 2003, 15.

[12] K.R. Popper, *Vermutungen und Widerlegungen: das Wachstum der wissenschaftlichen Erkenntnis*, Tübingen, 2000; K.R. Popper, W. Warren Bartley, *Realism and the aim of science*, London/New York, 1993.

Also, the comparative method will be applied. In particular, the study will concern the Italian and the US criminal law, to assess the extent to which the normative solutions, which have been adopted in such legal systems, correspond and/or differ.

The choice to compare the US and the Italian law is of special interest for at least two reasons: firstly, they represent the two main legal traditions in the Western World, namely civil law, and common law; secondly, from the procedural viewpoint, they both are adversarial systems.[13]

It is to be noted that the comparative law literature is reconsidering the contrasts between civil law and common law, as the traditional criteria of demarcation, which are based on the different relationship between the legislation (allegedly typical of the former) and the judicial decisions (allegedly typical of the latter) as sources of law,[14] turn out to be unfounded.[15] In fact, there are common law systems, like the US, in which the relevance of enacted law as a source has gradually increased; on the contrary, there are civil law systems, like Italy, in which the judicial precedents have become a fundamental source of law, especially the ones set by the Supreme Court (*Suprema Corte di Cassazione*) and the Constitutional Court, as well as the ones set by the European Court of Human Rights and by the European Court of Justice[16] (which gave rise to a phenomenon of 'cross-fertilization' between the legal tradition of their respective Member States).[17] Furthermore, both in common law and in civil law systems, the experience of constitutionalism implicates that the rule of law, the enshrinement of human rights, and essential values like non-discrimination represent, today, a shared normative heritage.[18]

[13] J.F. Nijboer, *The American Adversarial System in Criminal Cases: Between Ideology and Reality*, in *Cardozo Journal of International and Comparative Law*, 5, 1, 1997, 79-96; L.J. Fassler, *The Italian Penal Procedure Code: An Adversarial System of Criminal Procedure in Continental Europe*, in *Columbia Journal of Transnational Law*, 29, 1, 1991, 245-278; W.T. Pizzi, L. Marafioti, *The New Italian Code of Criminal Procedure: The Difficulties of Building an Adversarial Trial System on a Civil Law Foundation*, in *Yale Journal of International Law*, 17, 1, 1992, 1-40; M. Panzavolta, *Of hearsay and beyond: is the Italian criminal justice system an adversarial system?*, in *The International Journal of Human Rights*, 20, 5, 2016, 617-633.

[14] J. Dainow, *The Civil Law and the Common Law: Some Points of Comparison*, in *American Journal of Comparative Law,* 15, 3, 1967, 419-435.

[15] R. Sacco, *Legal Formants: A Dynamic Approach to Comparative Law (Installment I of II)*, in *The American Journal of Comparative Law,* 39, 1, 1991, 1-34; G. Grasso, *Politiche penali e ruolo della giurisprudenza: la sfida della legalità*, in C.E. Paliero, F. Viganò, F. Basile, G.G. Gatta (eds.), *La pena, ancora: fra attualità e tradizione: studi in onore di Emilio Dolcini*, Milano, 2018, 47-67; M. Donini, *Il diritto giurisprudenziale penale. Collisioni vere e apparenti con la legalità e sanzioni dell'illecito interpretativo*, in *Diritto penale contemporaneo*, 3, 2016, 22-38; M. Vogliotti, *Il giudice al tempo dello scontro tra paradigmi,* in *Diritto penale contemporaneo*, 2nd November 2016, https://archiviodpc.dirittopenaleuomo.org/d/5029-il-giudice-al-tempo-dello-scontro-tra-paradigmi (last access 20/03/2023).

[16] Y. Lupu, E. Voeten, *Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights,* in *British Journal of Political Science*, 42, 2, 2012, 413-439; M. Payandeh, *Precedents and Case-based Reasoning in the European Court of Justice*, in *International Journal of Constitutional Law*, 12, 3, 2014, 832-835.

[17] D. Rietiker, *Strange Bedfellows: The Cross-Fertilization of Human Rights and Arms Control: The European Court of Human Rights on Cases Involving Chemical Weapons and Anti-Personnel Mines*, in *Cyprus Human Rights Law Review*, 3, 2, 2014, 130-159; F.G. Jacobs, *Judicial Dialogue and the Cross-Fertilization of Legal Systems: The European Court of Justice,* in *Texas International Law Journal*, 38, 3, 2003, 547-556.

[18] H.B. Higgins, *The Rigid Constitution, in Political Science Quarterly*, 20, 2, 1905, 202-222; G. Sartori, *Constitutionalism: A Preliminary Discussion*, in *American Political Science Review*, 56, 4, 1962, 853-864; M.J.C. Vile, *Con-*

Nevertheless, differences, which sometimes are radical, exist and cannot be denied (for example, with regards to the theories of interpretation).[19] In summary, one can affirm that three elements coexist: *a)* a shared political and cultural substratum; *b)* a different technical and juridical substratum; and *c)* current questions that both the legal traditions do have in common. In general, all Western traditions are characterized, on the one hand, by the intelligibility of norms only where originating from procedures and institutions that are conceptually coordinated; on the other, by the fact that the legality is higher than sovereignty, that is, a legal order (above all, a constitutional order) cannot be overturned, even by politics.[20]

3. Literature Review

This article will consider and criticize the doctrine maintaining that AI systems may be considered as legal persons, that is, mainly the works by Gabriel Hallevy.[21] The opposite thesis will be argued, not only by reference to what has already been affirmed by numerous scholars,[22] but also by proposing further arguments based on the vitiated nature of the alleged parallel between the legal status of AI entities and of corporations.

As for the substantive criminal law essential concepts, some important works by Wayne R. LaFave and Austin W. Scott, Jr., George P. Fletcher, Jens David Ohlin, Jonathan Herring, and Paul H. Robinson are quoted.[23] Concerning the technical features of AI, I mainly refer to the well-known book *Artificial Intelligence: A Modern Approach* by Stuart J. Russell and Peter Norvig, as well as to some works by Paolo Traverso and Margaret A. Boden.[24]

stitutionalism and the separation of powers, Indianapolis, 1998; C.R. SUNSTEIN, *Designing democracy: what constitutions do,* Oxford, 2002; E. D'ORLANDO, F*undamental Rights and New European Constitutionalism: an Italian Approach,* in *Transition Studies Review*, 13, 1, 2006, 201-209.

[19] On this topic see amplius A. GAMBARO, R. SACCO, *Sistemi giuridici comparati*, Torino, 2018; K. ZWEIGERT, H. KÖTZ, *Introduction to comparative law*, Oxford/New York, 1998; M.D. DUBBER, T. HÖRNLE, *Criminal law: a comparative approach*, Oxford/New York, 2014.

[20] A. GAMBARO, R. SACCO, *op. cit.*, 31-45.

[21] G. HALLEVY, T*he Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4, 2, 2010, 171-201; ID., *Liability for Crimes Involving Artificial Intelligence Systems*, Cham, 2015.

[22] S. BECK, *Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, 86, 2016, 138-143; L. FLORIDI, *Digital's Cleaving Power and Its Consequences*, in *Philosophy & Technology*, 30, 2, 2017, 123-129; R. BROWNSWORD, *Law 3.0: rules, regulation and technology*, Abingdon/New York, 2020; F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto Penale e Uomo*, 10, 2019, 1-34.

[23] W.R. LAFAVE, A.W. SCOTT, *op. cit.*; G.P. FLETCHER, *Rethinking*, cit.; J.D. OHLIN, *op. cit.*; J. HERRING, *op. cit.*; P.H. ROBINSON, *Structure and function in criminal law*, Oxford/New York, 1997.

[24] S.J. RUSSELL, P. NORVIG, op. cit.; P. TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE Online*, 51, 1, 2022, 155-167; M.A. BODEN, *Artificial intelligence: a very short introduction*, Oxford, 2018.

4. Facts' Overview and Discussion

Agents or instruments? Legal subjects or mere objects? That is the question.[25]
When answering it, one should always consider the goals that are pursued by the way of law. In other words, normative solutions should be fit to achieve the ethical goals that have been set by the politics. Hence, solutions change depending on the fact that, in order of importance, the main ethical value is, for example, either the protection of human rights, or the economic and technological development whatever it takes.

4.1. Trustworthy AI

Both in the United States and in the European Union, the trustworthiness of AI and a responsible approach to such technology are among the policies to be emphasized.
As for the US, the *National AI Initiative Act* of 2020[26] (DIVISION E, SEC. 5001) became law on January 1, 2021. The mission of the National AI Initiative is to ensure continued US leadership in AI research and development, lead the world in the development and use of trustworthy AI in public and private sectors, and prepare the present and future US workforce for the integration of AI systems across all sectors of the economy and society. Also, the Executive Order 13859, issued by the President of the United States in February 2019, directed the Secretary of Commerce, through the National Institute of Standards and Technology (NIST), to issue "a plan for Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies".[27] Such plan, entitled *U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools*,[28] has been issued in August 2019. According to it, among the areas of focus for AI standards there is trustworthiness, meaning "guidance and requirements for accuracy, explainability, resiliency, safety, reliability, objectivity, and security". However, the plan clarifies that societal and ethical considerations in information technology (IT) consist of the analysis of its

[25] As for the Italian literature on this question, see *inter alia* M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, in A. CADOPPI, S. CANESTRARI, A. MANNA, M. PAPA (eds.), *Cybercrime - Diritto e procedura penale dell'informatica*, Torino, 2018, 1180-1212; A. FIORELLA, *Responsabilità penale del Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (eds.), *Il diritto nell'era digitale: Persona, Mercato, Amministrazione, Giustizia*, Milano, 2022, 651-664; S. MASSI, *Affidamento sull'intelligenza artificiale e "disimpegno morale" nella definizione dei presupposti della responsabilità penale*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (eds.), *Il diritto nell'era digitale*, cit., 665-680; D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (eds.), *Il diritto nell'era digitale*, cit., 681-694; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (eds.), *Il diritto nell'era digitale*, cit., 713-726; R. BORGONOVO, *La responsabilità penale nei processi ad elevata automazione*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (eds.), *Il diritto nell'era digitale*, cit., 727-744; B. GIULIANO, F. DE SIMONE, A. ESPOSITO, S. MANACORDA (eds.), *Diritto penale e intelligenza artificiale: nuovi scenari*, Torino, 2023; C. PIERGALLINI, I*ntelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Rivista italiana di diritto e procedura penale*, 4, 2020, 1743-1772.
[26] Available at https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210 (last access 01/11/2022).
[27] 84 FR 3967.
[28] Available at https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf (last access 01/11/2022).

nature and social impact, as well as the corresponding formulation and justification of policies for the appropriate use of it. Furthermore, in this regard, technical and non-technical standards should be distinguished, as not all societal and ethical issues of AI (for instance, in areas such as criminal justice and healthcare) can be addressed by developing technical standards, then organizations should develop AI systems that leverage human judgment and responsibility where they are needed.

As for the EU, *inter alia*, one can mention the *Ethics Guidelines for Trustworthy Artificial Intelligence* presented in August 2019[29] by the High-Level Expert Group on AI (AI HLEG), that is, an independent expert group that was set up by the European Commission in June 2018. According to such document, Trustworthy AI has three components, which should be met throughout the systems' entire lifecycle: *a)* lawful AI, namely, complying with all applicable laws and regulations; *b)* ethical AI, meaning that adherence to ethical principles and values shall be ensured; and *c)* robust AI, both from a technical and social perspective, because, even with good intentions, AI systems can cause unintentional harm. The foundations of Trustworthy AI are four ethical principles based on fundamental rights: *a)* respect for human autonomy; *b)* prevention of harm; *c)* fairness; and *d)* explicability. Given this, the realization of Trustworthy AI needs the implementation of the following key requirements, which shall be evaluated and addressed continuously throughout the systems' entire lifecycle, by means of both technical and non-technical methods: *a)* human agency and oversight; *b)* technical robustness and safety; *c)* privacy and data governance; *d)* transparency; *e)* diversity, non-discrimination, and fairness; *f)* societal and environmental wellbeing; and *g)* accountability.[30]

4.2. Automation versus Autonomy

One should wonder whether AIs are, in the proper sense, automatic or autonomous. As a matter of fact, the term "automatic" originates from the Ancient Greek adjective "*αὐτόματος*", which means "having a self-acting mechanism", and refers to entities that are capable to act spontaneously but by virtue of a deterministic relationship with another entity. Instead, the term "autonomous" derives from the adjective "*αὐτόνομος*", which means having a self-regulating mechanism, and refers to entities that are capable to act by virtue of their own 'laws' (*νομοί*): in other words, entities that are endowed with the freedom of the will and, therefore, with *mens rea*.

Therefore, either AIs are fully autonomous (in the proper sense), and then can be legal persons; or they are merely automatic (at the most, only partially autonomous), and can be considered as mere instruments solely.

[29] High-Level Expert Group on AI (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last access 01/11/2022).

[30] About the issue of trustworthy AI, see further European Commission, *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, COM(2021) 206 final, 2021/0106(COD), 21/04/2021, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (last access 20/03/2023).

ISSN 2284-4503

4.3. Thesis: AIs are legal subjects

Professor Gabriel Hallevy, quoting Professor Lawrence B. Solum,[31] affirmed that people are fearful of AIs because they "are not considered to be subject to the law, specifically to criminal law".[32] Such fear would resemble the one from which people suffered in the past with regards to corporations and their ability to commit crimes. However, according to the scholar, "because corporations are legal entities subject to criminal and corporate law, that kind of fear has been significantly reduced".[33]
He theorized three models of criminal liability involving AIs.

- In the perpetration-through-another model, AI entities are not considered as possessing any human attributes, but as innocent agents used by the actual perpetrator (principal), namely the programmer or the user. In other words, this model corresponds to the paradigm of perpetration-by-means, clearly described by Professor George P. Fletcher, in which the focus is on persons who are not actively engaged in the carrying out of the criminal act: for example, children or insane people used to implement a criminal plan of which they are totally unaware.[34]
- Similarly, in the second model, named natural-probable-consequence model, AI systems are deemed as objects, since the programmers and/or users can be held liable where an offense is committed via AI and occurs as a natural and probable consequence of their intentional or negligent conduct.
- Instead, the direct liability model "does not assume any dependence of the AI entity on a specific programmer or user"[35] and considers the AI entity as a legal person. In particular, while the fulfilment of the *actus reus*[36] "is easily attributed to AI entities", "[a]ttributing the internal element of offenses to AI entities is the real legal challenge".[37] According to the scholar, "[o]ne might assert that humans have feelings that cannot be imitated by AI software, not even by the most advanced

[31] L.B. SOLUM, *Legal Personhood for Artificial Intelligences*, in *North Carolina Law Review*, 70, 4, 1992, 1231-1287.
[32] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 173.
[33] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 174.
[34] G.P. FLETCHER, *Rethinking*, cit., 634-649.
[35] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 186.
[36] As for the structure of crimes, one should consider that Professor Paul H. Robinson proposed to innovate the traditional two-faced structure, that is, *actus reus* and *mens rea*. According to the scholar, such basic organizing distinction, "[r]ather being useful to criminal law theory, it is harmful because it creates ambiguity in discourse and hides important doctrinal differences of which criminal law should take account". On the one hand, actus reus usually include four kinds of requirements, namely "the conduct, circumstance, and results elements of an offence, as well as the supporting doctrines of causation, voluntary act, and omission and possession liability". On the other, *mens rea* "typically is said to be the actor's required mental state at the time of the conduct constituting the offence". However, "[w]hat is the unifying characteristic of the *actus reus* requirements? Are they all 'objective' in nature? ". The answer is negative, since an objective element can include a subjective state of mind, such as the case of negligence, which is "neither subjective nor a state of mind, of course, but rather a failure to meet an objective standard of attentiveness". Hence, in place of the traditional bipartite structure, Robinson proposes a tripartite structure of crimes, that is, an offense is composed of the following elements: *a)* the objective requirements, namely the conduct and, in the crimes within the pattern of harmful consequences, the occurrence of a result and the proof of causation; *b)* the culpability requirements (purpose, knowledge, recklessness, negligence); and *c)* the act-omission requirements, in case of legal duty to act (the commission-by-omission crimes). In this regard, see P.H. ROBINSON, *supra* note 23.
[37] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 187.

software", but "such feelings are rarely required in specific offenses" and "[m]ost specific offenses are satisfied by knowledge of the existence of the external element".[38] Thus, although traditionally only humans were subject to criminal law, since the English precedent of 1635 concerning the case of *Langforth Bridge* corporations have been treated as criminal law subjects.[39] Then, "[w]hy should AI entities be different from corporations? AI entities are taking larger and larger parts in human activities, as do corporations".[40]

4.4. Antithesis: AIs are mere objects

Hallevy's theory cannot be accepted, being it grounded upon fallacious arguments.

4.4.1. Is AI actually 'intelligent'?

AIs are not actually intelligent.

As Paolo Traverso recently explained, two main different approaches relate to AI, namely model-based and machine learning (ML)-based AI.

As for the model-based AI, the system imitates the behaviours of a given domain's experts. In a few words, the programmer, who is a computer science expert, hopefully with the help from experts in the fields each time considered (for example, medicine), defines the knowledge representation about a phenomenon (for example, myocardial infarction), and integrates such model into the system, so that the latter can 'treat' the phenomenon (for example, analysis of the risk-factor to which a patient is exposed, then calculation of the myocardial infarction's probability). There are two types of model-based AI.

- In systems that are based upon if-then rules, given the premise α, the system formulates the conclusion β, to solve the question γ that has been posed by the programmer.
- In systems that are based on the so-called 'trees,' the knowledge is organized by reference to a model that evokes the shape of an ideal tree, whose 'fronds' correspond to the different data-classification's alternatives, in such a way that, after the 'ramification,' an output is produced. To put it differently, the software recognizes the question γ by virtue of a series of variables, that is, given the starting situation α, the various possible alternatives 'ramify' until the conclusion β is reached.[41]

As for the ML-based AI, instead, a phenomenon's model is obtained from data that, for instance, are made available on the Internet, by the numerous sensors in our cities, as well as by the sensors with which our smart watches are equipped. It is a technique that allows to realize complex knowledge representations, by way of statistics and the probability theory. The ML is grounded on different learning methods and several computational models for data analysis.[42]

There are three ML learning methods.

[38] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 189.
[39] *Case of Langforth Bridge*, 79 Eng. Rep. 919 (K.B. 1635).
[40] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 200.
[41] P. TRAVERSO, *op. cit.,* 158-160.
[42] P. TRAVERSO, *op. cit*., 160, J.S. RUSSELL, P. NORVIG, *op. cit*., 704-713.

In the supervised learning, the programmer defines the so-called training sets, which include the data concerning a phenomenon, and consequently creates a computational model that is fit to 'apprehend' the information included in the training set (the correlations between the cases and the solutions). In other words, the programmer 'trains' the AI by defining a set of expected outcomes in relation to a certain input range, and by constantly evaluating the achievements of the objectives; then, the system formulates a hypothesis, and every time it makes a mistake, the hypothesis is reviewed.[43]

In the unsupervised learning, the programmer provides for neither expected results nor error-reports.[44] It is usually applied for the so-called 'clustering,' that is, the formation of sets that include elements presenting analogies or relevant connections (*e.g.*, documents concerning the same issue, individuals that have certain characteristics in common, as well as terms that serve the same purpose within a text).[45]

Finally, in the reinforcement learning, the system is led by a reward-punishment mechanism, that is, feedback messages regarding 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.[46]

Hence, it should be clear that unlike the model-based AI, in the ML the focus is not on the definition of a knowledge model, but on the collection of data and their inclusion within training set. Such data allow the training of the computational models, for example the 'artificial neural nets' (ANNs), namely neural computational systems that are inspired by the functioning of the human brain (the biological neural nets or BNNs). ANNs and BNNs present two similarities. On the one hand, the building blocks of both nets are highly interconnected computational tools. On the other, "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".[47] Then, "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".[48]

In Figure 1 we see a neural net. It is composed of some 'layers' (in fact parametric functions: functions in which the parameters are not defined) that are distributed in parallel. The task of the first layer of neurons (the blue circles) is to translate certain data into input signals. For instance, in the case of image recognition, data (in the form of numbers) represent the pixels' colours; such information is linked to other image's features; the various neurons are connected to each other, and the last layer of neurons (the red circles) does have the task to recognize the image's subject, so that, by virtue of

[43] P. Traverso, *op. cit*., 161.; J.S. Russell, P. Norvig, *op. cit.,* 653-656.

[44] F.C. La Vattiata, *Artificial Intelligence in Healthcare*, cit., 49.

[45] F. Lagioia, G. Sartor, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *Philosophy & Technology,* 33, 2020, 433-465; J.S. Russell, P. Norvig, *op. cit*., 775-581; M. Jafari, Y. Wang, A. Amiryousefi, J. Tang, *Unsupervised Learning and Multipartite Network Models: A Promising Approach for Understanding Traditional Medicine*, in *Frontiers in Pharmacology*, 11, 2020, 1319.

[46] F.C. La Vattiata, *Artificial Intelligence in Healthcare*, cit., 49; A. Jonsson, *Deep Reinforcement Learning in Medicine*, in *Kidney Diseases*, 5, 1, 2019, 18-22.

[47] F.C. La Vattiata, *Artificial Intelligence in Healthcare*, cit., 48-49.

[48] Y.-S. Park, S. Lek, *Artificial Neural Networks*, in *Developments in Environmental Modelling*, 28, 2016, 123-140.

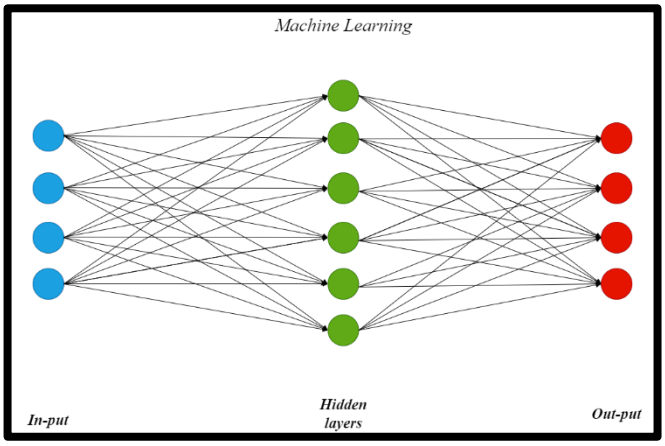
statistical analysis grounded upon the weight of each connection, the probability of the correspondence between the image and the aforementioned information can be calculated.

An interesting case of ML is deep learning (DL). As we can see in Figure 2, in DL the neural nets do have several intermediate 'hidden' layers of neurons (the green circles), and the programmer's training serves to define such parameters.[49]

For instance, by way of a technique called 'back-propagation' (or 'backprop'), the programmers 'teach' the output's object to the AI system (*e.g.*, in the case mentioned above, the image's subject that the system will have to recognize): for this purpose, they define each intermediate layer's parameters backward, until the first level (the input) is reached. Also, in DL the functions corresponding to the intermediate neurons are non-linear, otherwise the various layers could be reduced into a unique layer.[50]

In short, even in case of ML and DL, AIs are based on statistical (albeit advanced) calculations. In other words, neither such entities are merely 'automatic,' nor they are endowed with the degree of autonomy that is sufficient to make them 'autonomous' (in the sense clarified above). Indeed, as Professor Luciano Floridi affirmed, an AI system is not actually intelligent: it is "a counterfactual: were a human to behave in that way, we would call that behaviour intelligent. It does not mean that it is intelligent".[51] Hence, we can call an AI system's behaviour 'intelligent,' but only inasmuch as a human being, who behaves in such a way, is taken as a term for comparison.

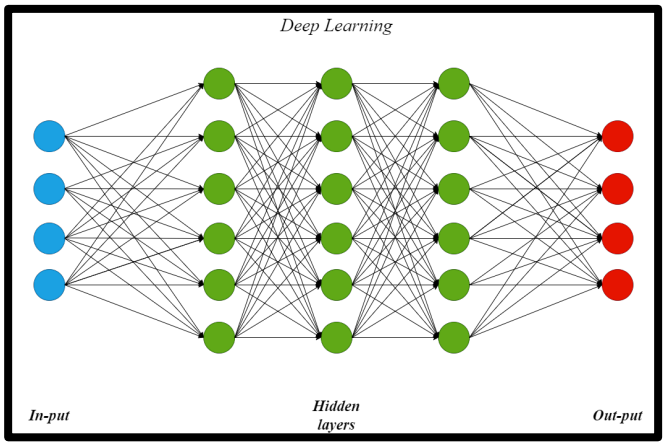
Figure 1 (Machine Learning)



[49] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*, Cambridge, 2016.
[50] P. Traverso, *op. cit.,* 163.
[51] L. Floridi, *op. cit.,* 91.

ISSN 2284-4503

Figure 2 (Deep Learning)



4.4.2. The substantial legal difference between corporations and AI systems

It is a false argument to affirm that there is "no substantial legal difference between the idea of criminal liability imposed on corporations and on AI entities",[52] since such substantial difference exists. Actually, while in his third model Hallevy assumes no dependence between the AI entity and a human being (either the programmer, or the user), the liability of corporations depends on a crime materially committed by a human being linked to the *societas*, both in the United States (so-called superior agent rule)[53] and in Italy (so-called *colpa di organizzazione*).[54]

In the US, a corporation is held liable for the conducts of its agents within the scope of their employment, provided that a "purpose to benefit the corporation is necessary to bring the agent's acts within the scope of his employment".[55] Such rule essentially specifies the principles regulating the so-called 'vicarious liability,' under which the offense (*i.e.*, both the *actus reus* and the *mens rea*), committed by a human being representing the corporation, is automatically attributed to the *societas*.[56]

In Italy, according to Law (*Decreto legislativo*)[57] no. 231 of 2001, the grounds for recognizing a corporation as responsible for a crime are the following.

[52] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 201.

[53] W.R. LAFAVE, A.W. SCOTT, *op. cit.*, 360.

[54] G. LOSAPPIO, *Organizzazione, colpa e sicurezza sul lavoro. Dosimetria dell'impresa e della colpa di organizzazione,* in *Diritto della sicurezza sul lavoro*, 2016, 98-112.

[55] *United States v. Hilton Hotels Corp.*, 467 F.2d 1000 (9th Cir. 1973).

[56] C. DE MAGLIE, *L'etica e il mercato: la responsabilità penale delle società*, Milano, 2002, 16; S. VINCIGUERRA, *Diritto penale inglese comparato: i principi,* 2nd ed., Padova, 2002, 278; about the concept of 'vicarious liability' see further G.P. FLETCHER, *Rethinking*, cit., 647-649.

[57] Pursuant to Article 76 of the Constitution, a *decreto legislativo* is a type of delegated legislation, that is, a legal act enacted by the Government, once authorized by the Parliament.

First, a natural person linked to the entity (a senior manager[58] or a subordinate employee[59]) shall commit one of the offenses listed in the Articles from 24 to 25-*duodevicies*.

Second, there shall be an objective link between the corporation and the perpetration of the crime, that is, in the first place, the offense shall be committed in the interest/for the benefit of the corporation[60] (Article 5, § 1). Anyway, the corporation cannot be held liable where the senior management or subordinate persons have acted in their own exclusive interest or in the exclusive interest of third parties (Article 5, § 2).

Third, one shall make a distinction depending on who is the perpetrator of the crime.

- Where the offense is committed by a senior manager (see Article 6), the corporation is not liable if it can demonstrate that: *a)* prior to the commission of the offense, a compliance programme aimed at preventing crimes of the type occurring has been adopted and has been efficiently enforced; *b)* the task of overseeing the corporation's activities and updating such models has been delegated to an organ (so-called *Organismo di Vigilanza*, hereinafter OdV) within the corporation; *c)* the OdV has been endowed with proper initiative and control powers; *d)* the superior agent committed the crime by fraudulently evading the compliance programme; and *e)* there has been no omission or insufficient oversight by the OdV.
- Where the offense is committed by a subordinate employee (see Article 7), the corporation is liable only in the case the commission of such offense has been made possible by virtue of noncompliance with the management or supervisory obligations. In any case, such noncompliance is excluded if the legal entity, prior to the commission of the offense, adopted and efficiently implemented a compliance programme which is adequate to prevent crimes of the same type as the one committed. The compliance programme shall provide for appropriate measures to ensure that the corporation's activity is carried out in compliance with the law, and to detect and eliminate risk situations in a timely manner, taking into account the organisation's nature and size and the type of activity carried out. The aforementioned compliance programme's effective implementation requires that: *a)* it shall be regularly assessed and (if need be) amended in case significant violations of its requirements are discovered, as well as changes occur in the organisation or in the business activity; and *b)* a disciplinary system adequate to sanction noncompliance with the programme's measures is provided for. [61]

[58] *I.e.,* persons holding a role of representation, administration and management in the legal entity or in one of its organisational units which is provided with financial and functional independence, as well as persons who, *de facto* or otherwise, manage and control the legal entity.

[59] *I.e.*, persons subject to the direction and supervision of the Senior Management.

[60] The interest of the corporation means that the perpetrator has acted with the purpose to help the legal entity, regardless of whether such objective has been reached. The benefit of the corporation means that it has achieved/could have achieved a positive result from the crime (even if non-economic).

[61] F.C. LA VATTIATA, *The prevention and punishment of corruption in the Italian legislation*, in *Revista Brasileira de Estudos Políticos,* 119, 2019, 117-147; V. MONGILLO, *La responsabilità penale tra individuo ed ente collettivo*, Torino, 2018.

4.4.3. *Machina delinquere et puniri non potest*

Among the fundamental principles regulating the criminal matter (at least in modern liberal-democratic societies), the well-known 'culpability' principle is worth being mentioned. It is sometimes summarised by the Latin formula *'nullum crimen, nulla poena sine culpa.'* In short, criminal law is conceived as an instrument for the protection of fundamental legal interests, such as life. Therefore, criminal punishment assumes the role of necessary instrument to make such a protection effective, on the one hand dissuading the people from the commission of crimes (so-called general prevention), and, on the other hand, avoiding the recidivism of those who have already been recognised as perpetrators of criminal offences (so-called special prevention). People need to self-determine their future behaviours, then the mental aspect is a necessary element of justification (albeit not the only one) for the punishment's infliction. If it were not, the deterrent function would certainly not be fulfilled, inflicting sanctions on persons who, when performed a certain conduct, were not *in dolo* or, at least, *in culpa*.[62] Hence, the relationship between the culpability principle and the functions attributable to punishment needs to be stressed.

Besides, from a broader perspective, both in the Anglo-Saxon and in the Italian theory of crime, for an offense to be fulfilled, the literature stresses the need for an act which is (at least potentially) 'voluntary.' In this regard, Section 2.01 of the US Model Penal Code can be mentioned: "(1) A person is not guilty of an offense unless his liability is based on conduct that includes a voluntary act or the omission to perform an act of which he is physically capable".[63] Also, as for the Italian legal system, the well-known concept of *'suitas'* is worth being mentioned, *i.e.*, the elements of 'consciousness' (*coscienza*) and 'will' (*volontà*) provided for in Article 42(1) of the Criminal Code, which summarise the conditions under which a conduct can be attributed to the actor. In other words, 'consciousness' and 'will' are the basic conditions under which an action or omission can be considered 'human' inasmuch it is encompassed within the 'domain of will' and, as a consequence, it can be differentiated from natural events as well as from merely mechanical inertia.[64] Abstractly, such a principle not only concerns cases of culpable liability, but also cases of objective liability, that is, cases for which the law does not require neither *dolus* nor *culpa*. Indeed, the Latin word *'suitas'* can be literally translated into English by reference to the concept of 'belonging,' meaning that a behaviour 'belongs' (= can be attributed) to the agent because of her/his 'consciousness and will,' which can be either potential in the case of *culpa* (hypothetical-normative element), or actual in the case of *dolus* (naturalistic-psychological element). In fact, a behaviour can be attributed to (= 'belongs' to) the agent – and accordingly deserves punishment – not only when it derives from a conscious impulse of the will (*dolus*), but also when it turns out to be avoidable through an effort of the will (*culpa*).[65]

[62] G. FIANDACA, *Considerazioni su colpevolezza e prevenzione*, in *Rivista italiana di diritto e procedura penale*, 4, 1987, 836-880.

[63] Also, according to P.H. ROBINSON, *Structure and function*, cit., 17: "[t]o summarize, actus reus is said to include the conduct, circumstance, and result elements of an offence, as well as the supporting doctrines of causation, voluntary act, and omission and possession liability".

[64] On the difference between 'human causes' and 'natural events' see further G.P. FLETCHER, *Basic concepts of criminal law*, New York, 1998, 59-73.

[65] In this regard, see amplius F. MANTOVANI, *Diritto penale: parte generale*, 11th ed., Milano, 2020, 327-331.

With this in mind, the question about the possibility to conceive AIs as possible authors of crimes and to inflict criminal sanctions upon them can be solved. By investigating the technical features of AIs, as we have already clarified (see *supra* § 4.4.1), even if (on the one hand) they are not merely automatic entities, they are not (on the other) endowed with the degree of autonomy that is sufficient to make them autonomous, meaning they are not endowed with the 'free will' which is a requisite of the 'capacity to be culpable.' By reference to the aforementioned elucidation by Professor Floridi – *i.e.*, an AI's behaviour can be called 'intelligent' only inasmuch as a human being behaving in such a way is taken as a term for comparison –, we can conclude that the law could hypothetically consider AIs 'intelligent,' however this would be a *fictio iuris* solely, whose opportunity from the criminal policy viewpoint is (at least) questionable.

In any case, even if AIs were able to commit crimes, they would not be able to be subject to punishment. That is to say, even if *machina delinquere potest*, surely *machina puniri non potest*.

In fact, should we admitted AI capable to directly commit a crime, then we would accept a nonsense: the criminal punishment would lose its functions. In other words, an indefectible element of all liability model would be lacking, that is, a subject to whom the burden of sanctions is to be allocated. In this regard, Professor Fabio Basile correctly argues that, inasmuch as we do not accept science fiction scenarios like digital criminal provisions understandable by the 'robotic community,' it is hard to theorize, with reference to AI entities, general prevention (*alias* general deterrence), that is, by punishing the perpetrators, the general public, on learning of the punishment, will be deterred from committing crimes.

If anything, one might assert that, among the theories of punishment,[66] the following are plausible: *a)* retribution, that is, the infliction of sufferings on the perpetrators by means of punishment aims at obtaining 'revenge' or, less emotionally, retribution for the harm they have caused to the victims in particular, and society in general; and *b)* special prevention, that is, by punishing the perpetrators they will be deterred from committing a crime again, as well as re-educated (or rehabilitated). For example, 'sanctions' like switching the 'liable' device off or imposing re-educational training on it could be provided for.[67]

However, such eventuality can be avoided through pragmatic measures, instead of theoretical and unrealistic solutions. Indeed, according to Professor Roger Brownsword, when challenges posed by the technical-scientific development are at stake, we should rely not only upon the legal rules' update and revision "so that they are fit to serve their intended purposes or policies", but also on "'technical' or 'technological' solutions", namely "a broad range of measures that might supplement or supplant the rules".[68] Specifically, such measures might be: *a)* 'architectural', that is, buildings, spaces and settings in general are re-designed or *ex novo* designed in a way that allows to manage the risks associated with certain technologies; *b)* incorporated in the design of products or processes, so that human beings

[66] Regarding the various theories of punishment, see J. HERRING, *op. cit.,* 62-66; W.R. LAFAVE, A.W. SCOTT, *Substantive criminal law,* 1, St. Paul, 1986, 30-41; G.P. FLETCHER, *Rethinking*, cit., 408-420.

[67] F. BASILE, *op. cit.,* 31-32.

[68] R. BROWNSWORD, *op. cit.,* 2.

ISSN 2284-4503

can be removed from potentially risky situations; or *c)* incorporated in wearables or even in humans themselves.[69]

Therefore, as an illustrative yet non exhaustive example, 'faulty conducts' of AI systems can be corrected either by ML algorithms which gradually optimize their behaviours, or by radically reprogramming them.[70]

5. Recommendations and Conclusive Remarks

In conclusion, we can affirm that AI entities should be considered as instruments, not as legal subjects. Such conclusion is surely valid from a *de iure condito* viewpoint, namely with regards to the law presently in force. However, it should be considered valid also from a *de iure condendo* viewpoint, that is, with respect to the law in a transitional stage, in the process of being established, or that is to be proposed.

In fact, the choice to consider the AI systems to be legal subjects would involve the risk of reducing human responsibilities, both legally and ethically. Consequently, the aforementioned strategy of Trustworthy AI and of a responsible approach to such technology would be unrealizable. Indeed, a 'slippery slope' to be avoided.

[69] R. BROWNSWORD, *op. cit.,* 2.

[70] A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018, 499-520.