

SP

SISTEMA
PENALE

FASCICOLO

2/2024

COMITATO EDITORIALE Giuseppe Amarelli, Roberto Bartoli, Hervè Belluta, Michele Caianiello, Massimo Ceresa-Gastaldo, Adolfo Ceretti, Cristiano Cupelli, Francesco D'Alessandro, Angela Della Bella, Gian Paolo Demuro, Emilio Dolcini, Novella Galantini, Mitja Gialuz, Glauco Giostra, Antonio Gullo, Stefano Manacorda, Vittorio Manes, Luca Maserà, Anna Maria Maugeri, Melissa Miedico, Vincenzo Mongillo, Francesco Mucciarelli, Claudia Pecorella, Marco Pelissero, Lucia Riscato, Marco Scoletta, Carlo Sotis, Costantino Visconti

COMITATO SCIENTIFICO (REVISORI) Alberto Alessandri, Silvia Allegrezza, Chiara Amalfitano, Ennio Amodio, Gastone Andrezza, Ercole Aprile, Giuliano Balbi, Marta Bargis, Fabio Basile, Alessandra Bassi, Teresa Bene, Carlo Benussi, Alessandro Bernardi, Marta Bertolino, Francesca Biondi, Rocco Blaiotta, Manfredi Bontempelli, Renato Bricchetti, David Brunelli, Carlo Brusco, Silvia Buzzelli, Alberto Cadoppi, Lucio Camaldo, Stefano Canestrari, Giovanni Canzio, Francesco Caprioli, Matteo Caputo, Fabio Salvatore Cassibba, Donato Castronuovo, Elena Maria Catalano, Mauro Catenacci, Antonio Cavaliere, Francesco Centonze, Federico Consulich, Stefano Corbetta, Roberto Cornelli, Fabrizio D'Arcangelo, Marcello Daniele, Gaetano De Amicis, Cristina De Maglie, Alberto De Vita, Ombretta Di Giovine, Gabriella Di Paolo, Giandomenico Dodaro, Massimo Donini, Salvatore Dovere, Tomaso Emilio Epidendio, Luciano Eusebi, Riccardo Ferrante, Giovanni Fiandaca, Giorgio Fidelbo, Carlo Fiorio, Roberto Flor, Luigi Foffani, Désirée Fondaroli, Gabriele Fornasari, Gabrio Forti, Piero Gaeta, Alessandra Galluccio, Marco Gambardella, Alberto Gargani, Loredana Garlati, Giovanni Grasso, Giulio Illuminati, Gaetano Insolera, Roberto E. Kostoris, Sergio Lorusso, Ernesto Lupo, Raffaello Magi, Vincenzo Maiello, Grazia Mannozi, Marco Mantovani, Marco Mantovani, Luca Marafioti, Enrico Marzaduri, Maria Novella Masullo, Oliviero Mazza, Claudia Mazzucato, Alessandro Melchionda, Chantal Meloni, Vincenzo Militello, Andrea Montagni, Gaetana Morgante, Lorenzo Natali, Renzo Orlandi, Luigi Orsi, Francesco Palazzo, Carlo Enrico Paliero, Lucia Parlato, Annamaria Peccioli, Chiara Perini, Carlo Piergallini, Paolo Pisa, Luca Pistorelli, Daniele Piva, Oreste Pollicino, Domenico Pulitanò, Serena Quattrocchio, Tommaso Rafaraci, Paolo Renon, Maurizio Romanelli, Gioacchino Romeo, Alessandra Rossi, Carlo Ruga Riva, Francesca Ruggieri, Elisa Scaroina, Laura Scomparin, Nicola Selvaggi, Sergio Seminara, Paola Severino, Rosaria Sicurella, Piero Silvestri, Fabrizio Siracusano, Nicola Triggiani, Andrea Francesco Tripodi, Giulio Ubertis, Maria Chiara Ubiali, Antonio Vallini, Gianluca Varraso, Vito Velluzzi, Paolo Veneziani, Francesco Viganò, Daniela Viganò, Francesco Zacchè, Stefano Zirulia

REDAZIONE Francesco Lazzeri, Giulia Mentasti (coordinatori), Enrico Andolfatto, Enrico Basile, Silvia Bernardi, Carlo Bray, Pietro Chiaraviglio, Stefano Finocchiaro, Beatrice Fragasso, Cecilia Pagella, Tommaso Trincherà

Sistema penale (SP) è una rivista *online*, aggiornata quotidianamente e fascicolata mensilmente, ad accesso libero, pubblicata dal 18 novembre 2019.

La *Rivista*, realizzata con la collaborazione scientifica dell'Università degli Studi di Milano e dell'Università Bocconi di Milano, è edita da Progetto giustizia penale, associazione senza fine di lucro con sede presso il Dipartimento di Scienze Giuridiche "C. Beccaria" dell'Università degli Studi di Milano, dove pure hanno sede la direzione e la redazione centrale. Tutte le collaborazioni organizzative ed editoriali sono a titolo gratuito e agli autori non sono imposti costi di elaborazione e pubblicazione.

La *Rivista* si uniforma agli standard internazionali definiti dal *Committee on Publication Ethics* (COPE) e fa proprie le relative linee guida.

I materiali pubblicati su *Sistema Penale* sono oggetto di licenza CC BY-NC-ND 4.00 International. Il lettore può riprodurli e condividerli, in tutto o in parte, con ogni mezzo di comunicazione e segnalazione anche tramite collegamento ipertestuale, con qualsiasi mezzo, supporto e formato, per qualsiasi scopo lecito e non commerciale, conservando l'indicazione del nome dell'autore, del titolo del contributo, della fonte, del logo e del formato grafico originale (salve le modifiche tecnicamente indispensabili). La licenza è consultabile su <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Peer review I contributi che la direzione ritiene di destinare alla sezione "Articoli" del fascicolo mensile sono inviati a un revisore, individuato secondo criteri di rotazione tra i membri del Comitato scientifico, composto da esperti esterni alla direzione e al comitato editoriale. La scelta del revisore è effettuata garantendo l'assenza di conflitti di interesse. I contributi sono inviati ai revisori in forma anonima. La direzione, tramite la redazione, comunica all'autore l'esito della valutazione, garantendo l'anonimato dei revisori. Se la valutazione è positiva, il contributo è pubblicato. Se il revisore raccomanda modifiche, il contributo è pubblicato previa revisione dell'autore, in base ai commenti ricevuti, e verifica del loro accoglimento da parte della direzione. Il contributo non è pubblicato se il revisore esprime parere negativo alla pubblicazione. La direzione si riserva la facoltà di pubblicare nella sezione "Altri contributi" una selezione di contributi diversi dagli articoli, non previamente sottoposti alla procedura di *peer review*. Di ciò è data notizia nella prima pagina della relativa sezione.

Di tutte le operazioni compiute nella procedura di *peer review* è conservata idonea documentazione presso la redazione.

Modalità di citazione Per la citazione dei contributi presenti nei fascicoli di *Sistema penale*, si consiglia di utilizzare la forma di seguito esemplificata: N. COGNOME, *Titolo del contributo*, in *Sist. pen. (o SP)*, 1/2023, p. 5 ss.

IL DIBATTITO SULLA RESPONSABILITÀ PENALE DIRETTA DELLE IA: “MOLTO RUMORE PER NULLA”?(*)

di Marco Edgardo Florio

Il contributo affronta sinteticamente il dibattito teorico sviluppatosi negli ultimi anni attorno al tema della responsabilità penale diretta delle IA (con conseguente loro sottoposizione a “pena”), al fine di comprendere se esso sia motivato da preoccupazioni realmente fondate oppure no. A tal fine, vengono esaminati: a) sia i principali modelli teorici elaborati a sostegno di una siffatta forma di responsabilità per le IA; b) sia le considerazioni che dovrebbero verosimilmente portare a mantenere il focus del diritto (ora e per il futuro) permanentemente orientato su responsabilità civili e penali strettamente “umane”.

SOMMARIO: 1. Premessa. – 2. I tentativi di delineare una responsabilità penale diretta delle IA e le loro inevitabili criticità. – 3. Abbiamo davvero bisogno di una responsabilità penale diretta delle IA? – 4. Alcune osservazioni conclusive.

1. Premessa.

Droni che uccidono sulle strade urbane; auto senza conducente che causano danni a cose o persone; *software* che svolgono, in collaborazione o addirittura in sostituzione dell'uomo, compiti sempre più sofisticati, a volte arrivando ad interferire negativamente con le condotte umane: se si riscontra un reato in questi incidenti, chi dovrà esserne ritenuto penalmente responsabile? Il programmatore del software, il suo produttore o l'utente finale? E siamo sicuri che in queste ipotesi, e in altre consimili, il sistema d'intelligenza artificiale (IA)¹ sia sempre e solo uno strumento inanimato nelle mani del vero responsabile (umano) del crimine? Oppure, in considerazione delle

(*) Il presente contributo riproduce in italiano (con qualche adattamento) il testo della relazione che l'autore ha svolto nel corso della 1st UNA Europa Cybercrime Conference “Exploring the Future of Crime” – LIMA Blog Launch Event (evento tenutosi in data 5-6 ottobre 2023, presso la KU di Leuven), già in corso di pubblicazione in inglese su: www.limablog.org.

¹ Il concetto di IA è tutt'altro che univoco (cfr. J. KAPLAN, *Intelligenza artificiale. Guida al futuro prossimo*, LUISS University Press, Roma, 2018, pp. 1 ss.; M.C. SCHEAU – A.L. ARSENE – G. POPESCU, *Artificial Intelligence/Machine Learning Challenges and Evolution*, in *Int' J. Info. Sec. Cybercrime*, 7, 2018, pp. 11 ss.; M. LANZI, *Self-driving cars e responsabilità penale. La gestione del “rischio stradale” nell'era dell'intelligenza artificiale*, Giappichelli, Torino, 2023, pp. 13 ss.), ma non potremo addentrarci *funditus* in questioni definitorie nella presente sede.

caratteristiche che alcune sistemi di IA possiedono, dovrebbero essere, almeno a volte, essi stessi considerati direttamente come i veri autori del reato².

Fino a non molti anni fa, domande del genere sarebbero state considerate di un certo interesse soltanto per gli appassionati di opere di fantascienza. Oggi, considerando che siamo molto più vicini di un tempo alla “distopica” realtà descritta da molti di tali lavori, simili interrogativi sembrano assai meno “banali” anche per la comunità scientifica e accademica. Non sorprende, dunque, che le difficoltà di concepire un intervento del diritto penale volto a punire gli esseri umani nel rispetto dei principi fondamentali che lo governano, per ciò che le macchine sembrano talvolta fare autonomamente, abbiano portato negli ultimi anni gran parte della letteratura a chiedersi se non fosse giunto il momento d’iniziare a pensare a una responsabilità penale diretta delle macchine³.

La questione nasce (e sembra acquisire il proprio significato), come è facile immaginare, dal fatto che ormai non esistono solo macchine che fanno ciò che il loro progettista ha “insegna” loro, ovvero che richiedono (per funzionare) sempre un significativo e costante controllo umano, ma anche macchine intelligenti che possono certamente operare in modo autonomo, senza alcuna supervisione umana⁴, e che a volte apprendono persino da sole (attraverso il *machine learning*), divenendo così imprevedibili per i loro stessi creatori.

Il vero nocciolo della questione non riguarda quindi le macchine basate su “algoritmi espliciti” (in cui la logica di funzionamento è stata interamente definita da programmatori umani)⁵. E non riguarda nemmeno le macchine costruite in tutto o in

² Cfr., per questi stessi interrogativi, già F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in www.dirittopenaleuomo.org, 29.9.2019, pp. 24 ss.; M.B. MAGRO, *Decisione umana e decisione robotica. Un’ipotesi di responsabilità da procreazione robotica*, in www.lalegislazionepenale.eu, 10.5.2020, pp. 4 ss.

³ Cfr., *ex multis*, S. GLEß – T. WEIGEND, *Intelligente Agenten und das Strafrecht*, in *ZStW*, 2014, pp. 561 ss.; G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Berlin-New York, 2014, pp. 185 ss.; E. HILGENDORF, *Können Roboter schuldhaft handeln?*, in S. BECK (hrsg), *Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Nomos, Baden-Baden, 2012, pp. 119 ss.; M. SIMMLER – N. MARKWALDER, 2017, *Roboter in der Verantwortung? Zur Neuaufgabe der Debatte um den funktionalen Schuldbegriff*, in *ZStW*, pp. 20 ss.; S. ZIEMANN, *Wesen, seid’s gewesen? Zur Diskussion über ein Strafrecht für Maschinen*, in E. HILGENDORF – J.-P. GÜNTHER (hrsg), *Robotik und Gesetzgebung*, Nomos, Baden-Baden, 2013, pp. 183 ss.; F. BASILE, *Diritto penale e intelligenza artificiale*, in *Giur. it.*, 2019, pp. 69 ss.; U. PAGALLO – S. QUATTROCOLO, *The impact of AI on criminal law, and its twofold procedures*, in W. BARFIELD – U. PAGALLO (eds), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar, Cheltenham-Northampton, 2018, pp. 385 ss.; D. LIMA, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, in *South Carolina Law Review*, 2018, pp. 677 ss.; T. KING – N. AGGARWAL – M. TADDEO – L. FLORIDI, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *Science and Engineering Ethics*, 2019, *passim*; P. ASARO, *A body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in P. LIN – K. ABNEY – G.A. BEKEY (eds), *Robot Ethics*, Harvard University Press, Cambridge, 2012, pp. 169 ss.

⁴ Si possono concepire molti livelli di automazione, variabili di settore in settore (cfr. *amplius* I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 1, 2021, pp. 90 ss.).

⁵ Così anche G. SEHER, *Intelligente Agenten als „Personen“ im Strafrecht?*, in S. GLEß – K. SEELMANN (hrsg), *Intelligente Agenten und das Recht*, Nomos, Baden-Baden, 2016, p. 46.

parte su algoritmi “impliciti” (algoritmi, cioè, la cui logica è almeno in parte costruita attraverso l’apprendimento automatico)⁶, quando producono eventi che sono il risultato di: (a) una programmazione espressamente preordinata a tal fine; (b) evidenti errori di progettazione, programmazione e/o costruzione; (c) cattiva gestione da parte dell’utente umano (a condizione che vi sia un controllo umano significativo e/o un ridotto grado di autonomia dell’IA). In questi casi, la responsabilità di ciò che la macchina ha causato – ma sembra forse improprio persino parlare di una vera e propria “causazione” da parte della macchina, dal momento che l’evento realizzato è qui soprattutto il risultato della cattiva gestione dell’utente umano o di errori (o *input*) di progettazione, costruzione e programmazione a monte – potrebbe essere attribuita (a titolo di dolo o colpa, a seconda del caso) agli esseri umani “dietro la macchina”. Qualsiasi tentativo di sostenere il contrario non sarebbe altro che uno stratagemma per distogliere l’attenzione da coloro che dovrebbero essere considerati realmente responsabili dei danni causati.

Non che non possano sorgere interessanti problemi teorici e pratici anche in questi casi.

(a) In primo luogo, si potrebbe porre il problema di determinare il grado di colpa che dovrebbe esser reputato “equo” per poter attribuire la responsabilità penale all’umano “dietro la macchina” (ad es., al conducente di una autovettura semi-autonoma in caso di incidente causato anche da una sua negligenza e/o mancanza di controllo)⁷.

(b) Inoltre, c’è anche la questione di stabilire cosa dovremmo fare se in una situazione di dilemma morale la macchina è (o deve essere in ogni caso) programmata per danneggiare necessariamente un qualche interesse al fine di evitare danni maggiori, finendo col fare esattamente ciò che era stata programmata a fare⁸.

(c) È importante capire, poi, anche cosa dovrebbe accadere quando fattori esterni determinano malfunzionamenti che risultano in seguito forieri di danni non pianificati a terzi. Supponiamo, ad es., che il *Park Distance Control System*, a causa di una particolare

⁶ Ricorriamo qui alla distinzione tra algoritmi “impliciti” ed “espliciti” delineata da A. JEAN, *Les algorithmes. Font-ils la loi?*, Éditions de l’Observatoire, Paris, 2021, pp. 32 ss.

⁷ A patto che un tale controllo sussista (sul tema si v. M. LANZI, *Self-driving cars*, cit., pp. 172 ss. e pp. 198 ss.). Favorevole rispetto ad una limitazione alla sola “colpa grave”, 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 (a cura di), *Il diritto nell’era digitale. Persona, Mercato, Amministrazione, Giustizia*, Giuffrè, Milano, 2022, pp. 660 ss.

⁸ Il problema delle situazioni “dilemmatiche” ha suscitato particolare interesse in dottrina: A. ENGLÄNDER, *Das selbstfahrende Kraftfahrzeug und die Bewältigung dilemmatischer Situationen*, in *ZIS*, 9, 2016, pp. 609 ss.; A. HEVELKE – J. NIDA-RÜMELIN, *Selbstfahrende Autos und Trolley-Probleme: Zum Aufrechnen von Menschenleben im Falle unausweichlicher Unfälle*, in *Jahrbuch für Wissenschaft und Ethik*, 19, 2015, pp. 5 ss.; E. HILGENDORF, *Autonomes Fahren im Dilemma. Überlegungen zur moralischen und rechtlichen Behandlung von selbsttätigen Kollisionsvermeidungssystemen*, in E. HILGENDORF (hrsg), *Autonome Systeme und neue Mobilität*, Nomos, Baden-Baden, 2017, p. 143; J.C. JOERDEN, *Zum Einsatz von Algorithmen in Notstandslagen. Das Notstandsdilemma bei selbstfahrenden Kraftfahrzeugen als strafrechtliches Grundlagenproblem*, *ivi*, pp. 73 ss.; G.M. SANDER – J. HOLLERING, *Strafrechtliche Verantwortlichkeit im Zusammenhang mit automatisiertem Fahren*, in *NStZ*, 2017, pp. 193 ss.; C. ROXIN – L. GRECO, *Strafrecht. Allgemeiner Teil*⁴, I, C.H. Beck, München, 2020, Rn. 125d ss.; I. COCA VILA, *Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law*, in *Criminal Law & Philosophy*, 1, 2018, pp. 59 ss.; M. LANZI, *Self-driving cars*, cit., pp. 48 ss.; O. DI GIOVINE, *Dilemmi morali e diritto penale. Istruzioni per un uso giuridico delle emozioni*, Il Mulino, Bologna, 2022, *passim*.

condizione climatica, non rilevi la presenza di un bambino dietro l'autovettura, e che quindi il veicolo autonomo urti quest'ultimo nella manovra di parcheggio. Dovremmo considerare generalmente responsabile l'utente, come la giurisprudenza ha stabilito per gli incidenti causati dai sistemi di controllo della distanza di parcheggio in Germania? Oppure, se l'evento era imprevedibile, dovremmo seguire l'opinione di Beck, secondo cui non sarebbe «convincente che l'utente sia» considerato «pienamente responsabile solo per la sua decisione di utilizzare la macchina, poiché ciò renderebbe le macchine inutili sotto molti punti di vista» (*enfasi aggiunta*)⁹? E perché non addossare invece la responsabilità di tutti i danni al produttore dell'IA?

Tutti questi problemi teorici (così come altri), tuttavia, non sembrano richiedere necessariamente di arrivare a concepire una responsabilità penale diretta delle macchine per poter essere risolti, dal momento che possiamo sempre intravedere un "centro di imputazione umana" dietro di esse. Che poi si decida d'imputare sempre penalmente i danni causati allo stesso è tutta un'altra storia.

Piuttosto, la questione dell'eventuale necessità di una responsabilità penale diretta delle macchine sembra esser principalmente sollecitata – presumiamo – da quelle macchine che si affidano ad algoritmi in tutto o in parte "impliciti", quando però agiscono in piena (o comunque con un elevato livello di) "autonomia" e producono danni che sono il risultato dei processi di elaborazione "imperscrutabili" che sono alla base della capacità di apprendimento loro conferita. In questo caso, infatti, oltre ai problemi che potrebbero sorgere dal punto di vista della determinazione del nesso di causalità¹⁰, imputare penalmente a titolo di colpa ogni evento che sia causato (contravvenendo o andando al di là dalla programmazione iniziale) da una IA intrinsecamente "imprevedibile" e del tutto autonoma, sia all'utente finale sia a chi l'abbia programmata o prodotta¹¹, darebbe probabilmente luogo a una responsabilità oggettiva pura, uno sfregio significativo al volto del diritto penale, in aperto contrasto con i principi garantistici che dovrebbero sempre governarlo.

Come è noto, infatti, nei sistemi di *civil law* il diritto penale è generalmente retto dal c.d. principio di colpevolezza. Questo porta solitamente a un radicale divieto di responsabilità oggettiva. Nei sistemi di *common law* la situazione è leggermente diversa, poiché la responsabilità oggettiva ha sempre avuto un ruolo importante anche nel diritto

⁹ Così S. BECK, *Robotics and Criminal Law. Negligence, Diffusion of Liability and Electronic Personhood*, in E. HILGENDORF – J. FELDLE (eds), *Digitalization and the Law*, Nomos, Baden-Baden, 2018, p. 53; ma cfr. anche G. QUINTERO OLIVARES, *La robótica ante el derecho penal. El vacío de respuesta jurídica a las desviaciones incontroladas*, in *Revista Electrónica de Estudios Penales y de la Seguridad*, 1, 2017, pp. 16 ss. il quale ritiene che la "casualità" dell'evento per l'utente finale debba esser riconosciuta in questi casi.

¹⁰ Per un primo inquadramento delle questioni che potrebbero porsi su tale fronte, cfr., nella dottrina italiana, D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, in R. GIORDANO – A. PANZAROLA – A. POLICE, S. PREZIOSI – M. PROTO (a cura di), *Il diritto nell'era digitale*, cit., p. 686.

¹¹ Sulle difficoltà riscontrabili nel riconoscere una responsabilità del produttore, cfr. B. FRAGASSO, [La responsabilità penale del produttore di sistemi di intelligenza artificiale](#), in *Dir. pen. cont. – Riv. trim.*, 1, 2023, pp. 26 ss.; M. LANZI, *Self-driving cars*, cit., pp. 205 ss. I "difetti" dell'apprendimento automatico, infatti, rientrerebbero verosimilmente nei "difetti da rischio di sviluppo" descritti da C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, Giuffrè, Milano, 2004, pp. 46 ss.

penale. Pure negli ordinamenti di *common law*, tuttavia, almeno negli ultimi anni, la responsabilità oggettiva ha iniziato ad esporsi a varie critiche e ad un progressivo “ridimensionamento”¹². Sicché, ad oggi, non è da escludere che un’imputazione per responsabilità oggettiva non possa comunque finire per dar luogo a qualche problema pure in questi ultimi sistemi, e non già soltanto nei primi.

Una imputazione a titolo di colpa, tuttavia, potrebbe talvolta persino apparire insufficiente agli occhi della collettività, facendo così emergere un *liability gap*, poiché si potrebbe arrivare a punire per colpa (nella migliore delle ipotesi) ciò che sembra invece attribuibile a una “malevola” volontà umana.

(a) Si pensi a situazioni in cui, ad es., una persona programma dolosamente un *robot* per commettere uno o più reati specifici, ma lo stesso finisce (o inizia nel tempo) a commettere (al posto o in aggiunta a quelli programmati) altri reati¹³.

(b) Si pensi poi a quei casi in cui il programmatore si limita a definire il *genus* del reato (ad es. l’omicidio) senza definirne altresì il *tempus* e il *modus*, lasciandoli alla discrezionalità della macchina¹⁴.

Anche per quanto riguarda questi ultimi casi, tuttavia, si possono sollevare fondati dubbi sulla divisibilità dei tentativi di elaborare una responsabilità penale diretta delle macchine. Tentativi che, forse più di quanto avrebbero dovuto, hanno negli ultimi anni catalizzato il dibattito dottrinale all’interno di molteplici sistemi giuridici.

2. I tentativi di delineare una responsabilità penale diretta delle IA e le loro inevitabili criticità.

I tentativi di delineare una nuova forma di responsabilità penale diretta per le macchine hanno seguito molte direzioni e dato origine ad una varietà di “modelli di giustificazione”. Ne riassumeremo brevemente alcuni, prima di delinearne gli aspetti maggiormente critici.

(a) Un primo orientamento fa leva sulla possibilità di riconoscere una personalità giuridica alle c.d. macchine “pensanti” (meglio: una personalità elettronica, giuridicamente rilevante), che consentirebbe di vedere nella macchina un autonomo

¹² Cfr. R.A. DUFF, *The Realm of Criminal Law*, Oxford University Press, Oxford, 2018, pp. 1 ss.; W.R. LAFAVE, *Substantive Criminal Law*³, Eagan, Minnesota, 2018, p. 572; M.D. DUBBER, *An introduction to the Model Penal Code*², Oxford University Press, Oxford, 2015, p. 65, evidenziando che nel *Model Penal Code* questa forma di responsabilità copre ormai solo ipotesi minori, «è un’opzione solo per gli illeciti civili *sui generis* previsti dal codice, le “violazioni”».

¹³ Cfr. S.-W. LEE, *Can an Artificial Intelligence Commit a Crime?*, in A. BRUNS – Y. BU – S. MEIER – H. MERKT – M. PAWLIK – E. TAKAHASHI – S. VÖNEKY – J. VON HEIN (eds), *Legal Theory and Interpretation in a Dynamic Society*, Nomos, Baden-Baden, 2021, p. 318; G. HALLEVY, *When Robots Kill. Artificial Intelligence under Criminal Law*, Northeastern University Press, Boston, 2013, p. 75.

¹⁴ In Italia gli studiosi ritengono che «in assenza di prova di istruzioni illecite che precisamente identifichino, anche dal punto di vista spazio-temporale, oltre che del *modus*, il» fatto criminoso «commissionato all’agente artificiale e da questi poi realizzato», non vi sarebbe «possibilità di condurre a processo il programmatore o l’utilizzatore del sistema di AI» con un’accusa di «dolo»: F. CONSULICH, *Flash Offenders. Le prospettive di Accountability penale nel contrasto alle intelligenze artificiali devianti*. In: *Riv. it. dir. proc. pen.*, 3, 2022, p. 1031.

centro d'interessi e di imputazione. Anche se concepita per operare principalmente sul piano della responsabilità civile (come avviene nelle ricostruzioni di alcuni autori, dove la macchina è destinata a rispondere dei danni causati attraverso un proprio fondo patrimoniale, alimentato *pro rata* da tutti i soggetti che hanno partecipato alla sua creazione, gestione e utilizzo), questa teoria contiene ovviamente in sé anche il germe di una potenziale e futura responsabilità penale diretta delle macchine. Non è infatti difficile immaginare che, seguendo la tesi, le macchine, in quanto persone giuridiche, potrebbero divenire presto o tardi anche destinatarie di misure interdittive, inabilitanti o d'adempimento (ad es., una misura di "disattivazione" del sistema quando provoca un danno), o essere considerate persino come le potenziali vittime di un reato¹⁵.

(b) Un secondo approccio, di stampo più strettamente funzionalista, parte dall'esame di concetti come il libero arbitrio e la soggettività penale, facendo particolarmente leva sulle tesi di Jakobs. Entrambi i concetti, in questa prospettiva, agirebbero come mere costruzioni interne al sistema sociale, indipendenti dalle caratteristiche biofisiche possedute dal soggetto di riferimento. La "persona", insomma, non sarebbe altro che un "costrutto" sociale, proiettabile in direzione di qualsiasi entità capace di deludere le aspettative del sistema. La concepibilità di una responsabilità penale diretta delle macchine dipenderebbe quindi, in ultima analisi, dal ruolo che la società stessa sarebbe disposta ad attribuire loro in futuro. Certamente nulla impedirebbe alle macchine, almeno in teoria, di divenire destinatarie dirette di sanzioni penali (multe o vere e proprie "pene" di "riprogrammazione", "conformazione" o "neutralizzazione"). Infatti, partendo da un concetto di responsabilità come prodotto dell'interazione sociale, la macchina non dovrebbe possedere una vera e propria *mens rea* per poter essere punita, ma una semplice *mindless morality*, che ben potrebbe fondarsi sulle sue caratteristiche d'interattività, adattabilità e autonomia¹⁶.

(c) Infine, il modello forse più compiuto rimane ancora oggi quello sviluppato da Hallevy¹⁷. Questa ricostruzione prende le mosse da un accostamento con il problema della responsabilità penale delle persone giuridiche¹⁸. Dal momento che la responsabilità

¹⁵ Cfr. S. RIONDATO, *Robotica e diritto penale (robot, ibridi, chimere, "animali tecnologici")*, in D. PROVOLO – S. RIONDATO – F. YENISEY (eds), *Genetics, Robotics, Law, Punishment*, Padova University Press, Padova, 2014, pp. 602 ss., secondo il quale, almeno in via ipotetica ed «entro certi limiti, alcuni raffinati soggetti di intelligenza artificiale potrebbero essere equiparati a "forme di vita" non-umane. Del resto, diversi sistemi di diritto penale tutelano già esseri non-umani da certe offese. L'esempio più evidente è quello della tutela penale degli animali (non-umani)».

¹⁶ Cfr. M. SIMMLER – N. MARKWALDER, *Rethinking the nature of culpability and legal personhood in an age of artificial intelligence*, in *Criminal Law Forum*, 30, 2019, pp. 1 ss.; D. PIVA, *Machina*, cit., pp. 690-691.

¹⁷ Cfr. G. HALLEVY, *I, Robot - I, Criminal - When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses*, in *Syracuse Sci. & Tech. L. Rep.*, 22, 2010, pp. 1 ss.; ID., *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4(2), 2010, pp. 171 ss.; ID., *Virtual Criminal Responsibility*, in *Original Law Review*, 6(1), 2010, pp. 6 ss.; ID., *When Robots Kill*, cit., *passim*; ID., *Liability for Crimes*, cit., *passim*.

¹⁸ Per un simile accostamento v. già L. SOLUM, *Legal Personhood for Artificial Intelligences*, in *NCLR*, 1992, 1248; più di recente, cfr. P.M. FREITAS, F. ANDRADE e P. NOVAIS, *Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible*, in P. Casanovas, U. Pagallo, M. Palmirani e G. Sartor (eds), *AI Approaches to the Complexity of Legal Systems. AICOL 2013 International Workshops*, Springer, Berlin-New York, 2014, 145 ss.

delle persone giuridiche ha trovato spazio in molti ordinamenti, dopo la “sepoltura” dei pregiudizi che si opponevano tenacemente alla sua affermazione, perché lo stesso non potrebbe accadere anche per le IA? Così come sono stati superati i dubbi sulla responsabilità penale degli enti, allo stesso modo dovrebbe superarsi l’idea che le macchine non possano essere rimproverate per i loro comportamenti dolosi o colposi, e di conseguenza punite. Le macchine più “intelligenti”, come gli esseri umani, sarebbero in grado di commettere atti assistiti dai necessari requisiti di intenzionalità e negligenza. Non ci sarebbe quindi motivo di astenersi dalla loro punizione, che potrebbe, a seconda dell’autore che sostiene la tesi, acquisire una funzione retributiva, specialpreventiva e forse persino generalpreventiva.

Secondo alcuni, infatti, «attraverso lo spegnimento definitivo o temporaneo della macchina, o attraverso la sottoposizione della macchina, dotata di congegni di autoapprendimento, ad un nuovo *training* “rieducativo”, si potrebbero realizzare, rispettivamente, la funzione retributiva e la funzione special-preventiva della pena»¹⁹. Altri, invece, sono contrari all’idea che si possa concepire una vera e propria “retribuzione” per le macchine²⁰ e spingono soprattutto sulle sanzioni incapacitanti e riabilitanti²¹. Maggiori dubbi, anche tra i sostenitori della responsabilità penale diretta dei sistemi di IA, gravitano invece attorno alla concepibilità o meno di una vera e propria deterrenza per medesimi²². L’idea di elaborare una responsabilità penale diretta per le macchine, tuttavia, qualunque strada si scelga di seguire, sembra andar inevitabilmente incontro ad alcuni ostacoli rilevanti.

Come anche è stato osservato, tra i sistemi di IA e le persone giuridiche ci sono senz’altro alcune analogie, ma anche molteplici differenze, che rendono difficile avallare l’idea di potersi limitare a trasporre *sic et simpliciter* gli argomenti spesi a sostegno della responsabilità penale delle seconde per giustificare anche quella dei primi²³. Una di queste differenze sembra acquisire una rilevanza fondamentale: diversamente dalle persone giuridiche, i sistemi di IA non sono costituiti da agenti umani, e questo fa la differenza nel concepire l’idea di poter applicare loro una pena. Le imprese collettive possono eventualmente derivare la loro colpevolezza dai loro attori umani, ai quali i precetti e le sanzioni sono fondamentalmente rivolti²⁴. Non hanno quindi un “corpo da

¹⁹ Cfr. F. BASILE, *Intelligenza artificiale e diritto penale*, cit., 2019, pp. 30-31; J. KAPLAN, *Intelligenza artificiale*, cit., p. 156.

²⁰ Cfr. HALLEVY, *When Robots Kill*, cit., p. 210, laddove osserva che «una macchina non soffre, e finché la retribuzione si baserà sul soffrire, la retribuzione non sarà molto rilevante per punire i robot».

²¹ Cfr. D. PIVA, *Machina*, cit., pp. 691 ss.

²² *Contra*, ad es., ritenendo che questa idea sia per il momento fondata su ipotesi vieppiù fantascientifiche, F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 32.

²³ Peraltro, non va dimenticato che ad oggi vi sono anche taluni dubbi in merito al fatto che la responsabilità penale degli enti possa ritenersi una responsabilità autenticamente “colpevole” (cfr. J.M. SILVA SÁNCHEZ, *Las personas jurídicas no pueden ser “culpables”. Un alegato para la creación de un subsistema propio*, in L. CORNACCHIA – E. DEMETRIO CRESPO (a cura di), *Responsabilità da reato degli enti collettivi Profili dogmatici e politico-criminali a oltre vent’anni dal d.lgs. 231/2001*, Giappichelli, Torino, 2023, pp. 125 ss.).

²⁴ Cfr. C. PIERGALLINI, *Intelligenza artificiale: da ‘mezzo’ ad ‘autore’ del reato?*, in *Riv. it. dir. proc. pen.*, 4, 2020, pp. 1768 ss, là dove osserva che «la *societas*, sia che la si interpreti in guisa di uno “schermo”, sia alla stregua di un “organo”, esiste nella realtà giuridica e sociale, ma è animata, naturalisticamente e spiritualmente, dagli

prendere a calci”, ma possono forse ricavare indirettamente la loro “anima da dannare” dai propri attori umani. Le IA, invece, non hanno alcun soggetto umano da cui attingere la propria “anima”, e questo ha conseguenze ragguardevoli per il discorso che stiamo affrontando. L’idea di una responsabilità penale delle macchine potrebbe soltanto implicare, infatti, un concetto “dimidiato” di azione e di colpevolezza, nonché (soprattutto) un concetto di “punizione” sostanzialmente privo di senso.

Dal primo punto di vista, in effetti, per poter concepire una responsabilità penale diretta delle IA, si dovrebbe verosimilmente far affidamento su una concezione “meccanicistica” di azione, ignorando le conquiste teoriche delle dottrine finalistiche post-welzeliane²⁵: per coloro che in materia di diritto penale sostengono concetti più ambiziosi di azione, in cui si presuppone una particella di “spirito” (*Geist*), una coscienza rudimentale o un qualche tipo di vita interiore, l’idea di una macchina che agisce sarebbe probabilmente considerata come una vera e propria *contradictio in terminis*²⁶. Inoltre, tale idea implicherebbe anche un concetto “vuoto” di colpevolezza²⁷, sprovvisto cioè di un qualsivoglia fondamento ontologico di carattere psichico²⁸.

Tuttavia, il fatto che molti interpreti attuali non avrebbero problemi ad accettare concetti ben più “modesti” di azione e/o di colpevolezza (che nel dibattito scientifico moderno stanno probabilmente divenendo persino predominanti) significa che non dovremmo limitarci a valutare soltanto queste critiche come potenziali ostacoli a una futura responsabilità penale diretta delle macchine. Il dibattito non può essere dissolto totalmente in questo modo. Il nocciolo della questione, infatti, va verosimilmente ricercato nel fatto che, anche se si concepisse che le macchine, prima o poi, potrebbero realmente arrivare ad agire ed essere colpevoli, la punizione inflitta loro non avrebbe comunque alcun significato²⁹.

In effetti, da questo punto di vista, l’essere umano colpito dalla pena dovrebbe essere concepito molto meno come un essere che “agisce” o “decide”, e più come un essere “unico”, “vulnerabile” ed “effimero”; decisivo non dovrebbe ritenersi tanto il fatto che egli possieda o no una coscienza o il libero arbitrio, quanto il fatto che possa

uomini che le danno vita: i precetti che le vengono rivolti sono diretti ai soggetti che l’hanno creata e che la rappresentano, condizionandone il comportamento (specie in vista della dotazione di un’adeguata organizzazione preventiva dei reati)»; S. GLEß – E. SILVERMAN – T. WEIGEND, *If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability*, in *New Crim. Law Rev.*, 19(3), 2016, p. 423.

²⁵ Sul tema cfr. S. GLEß – E. SILVERMAN – T. WEIGEND, *If Robots*, cit., pp. 419 ss.

²⁶ Per una difesa dell’idea che le macchine non possano “agire”, basandosi su una teoria del reato ancorata alla filosofia del linguaggio, si v. inoltre P.C. BUSATO, *De máquinas y series vivos: ¿Quién actúa en los resultados delictivos derivados de decisiones cibernéticas?*, in E. DEMETRIO CRESPO (a cura di), *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022, pp. 365 ss.

²⁷ Come quello adottato da M. SIMMLER – N. MARKWALDER, *Rethinking*, cit., pp. 1 ss.; D. PIVA, *Machina*, cit., 2022, pp. 690-691.

²⁸ Per una difesa dell’idea che non si possa concepire un’autentica colpevolezza delle IA, cfr. A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018, pp. 511 ss.; M.B. MAGRO, *Decisione umana*, cit., 10; S. GLEß, E. SILVERMAN e T. WEIGEND, *If Robots*, cit., pp. 421 ss.

²⁹ Cfr. C. ROXIN – L. GRECO, *Strafrecht*, cit., p. 370; S. GLEß – E. SILVERMAN – T. WEIGEND, *If Robots*, cit., pp. 423 ss.

essere consapevole della sua *conditio humana* come *humana fragilitas*³⁰. Tale richiamo, infatti, riporta alla mente anche ciò che di negativo c'è nella pena, la ragione per cui la tradizione del liberalismo giuridico-penale è riuscita ad ergere tante barriere e garanzie contro di essa: la pena colpisce gli esseri umani proprio negli aspetti centrali della loro *conditio humana*, toglie loro – come una pena detentiva – anni di una vita che possono vivere una volta sola, anni che andranno persi per sempre (siamo quindi molto più vicini alla posizione di Feinberg che a quella di Foot quando si tratta di stabilire che tipo di interessi la pena dovrebbe poter essere in grado di “comprimere”)³¹.

Come osserva Greco, «non ha senso parlare di punizione per i robot, perché la “punizione” che si può concepire» per questi ultimi, «senza la partecipazione della macchina alla *conditio humana* descritta», «sarebbe un *aliud* completo rispetto alla punizione reale». E l'argomentazione in parola rimarrebbe valida anche se la ricerca sulla c.d. «coscienza delle macchine» un giorno portasse a credere che le macchine siano realmente in grado di pensare, o se gli sforzi verso la c.d. “incarnazione” finissero per conferire alle macchine (cinque o più) sensi. Il «robot super intelligente del futuro, che in un certo senso può anche agire e decidere, è, in quanto sistema complesso di elaborazione di informazioni, anche copiabile». Ma qualcosa che può essere copiato non ha verosimilmente alcun motivo di temere per la propria esistenza – unica e irripetibile – e quindi non può essere «punito» in senso propriamente penalistico, ma solo “punito”. Tutto il resto che occorre affrontare quando si pensa al diritto penale dipende da questo³².

In sintesi: «la pena per i robot avrebbe solo il nome di pena, senza essere una pena» per davvero; ma se «ci possono essere» più modeste «alternative a concetti ambiziosi di azione e di colpevolezza», «non c'è alternativa» alcuna «a un concetto ambizioso di pena»³³. L'inflizione di una punizione ai sistemi di intelligenza artificiale non avrebbe quindi alcun senso e alcuna funzione da poter svolgere, a meno che non si sostenga un ritorno ad alcuni dei tempi più bui e irrazionali della storia umana³⁴, quando le persone erano solite “punire” gli animali o gli oggetti inanimati solo per soddisfare i

³⁰ La differenza ontologica che persiste tra umani e IA è riconosciuta da molti: E.J. LARSON, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, Harvard University Press, Cambridge, 2021, *passim*; N. CRISTIANINI, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Il Mulino, Bologna, 2023, *passim*; M. CHIRIATTI, *Incoscienza artificiale. Come fanno le macchine a prevedere per noi*, LUISS University Press, Roma, 2021, *passim*.

³¹ Cfr. J. FEINBERG, *The Rights of Animals and Unborn Generations*, in *Philosophy and Environmental Crisis*, 1974, pp. 49 ss.; P. FOOT, *Natural Goodness*, Clarendon Press, Oxford, 2001, *passim*; in argomento cfr. anche R. ABBOTT – A. SARCH, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, in *UC Davis Law Review*, 53(1), 2019, p. 365.

³² Cfr. L. GRECO, *Poder de julgar sem responsabilidade de julgador: a impossibilidade jurídica do juiz-robô*, Marcial Pons, São Paulo, 2020, p. 59.

³³ Cfr. L. GRECO, *Direito penal para robôs? Só se poderá falar em uma pena para robôs quando o direito penal deixar de se interessar pelo ser humano*, in *www.jota.info*, 5.11.2021, *passim*.

³⁴ Cfr. C. MULLIGAN, *Revenge Against Robots*, in *South Carolina Law Review*, 3, 2018, p. 578; *contra* K. DARLING, *Extending Legal Protection to Social Robots: The Effects of Antropomorphism, Emphaty, and Violent Behaviour Towards Robotic Objects*, in R. CALO – A.M. FROOMKIN – I. KERR (eds), *Robot Law*, Edward Elgar, Cheltenham-Northampton, 2016, p. 213.

loro impulsi vendicativi³⁵. Anziché un avanzamento per il diritto penale, l'attribuzione di una responsabilità penale diretta alle macchine, con la loro conseguente "punizione", porterebbe in definitiva soltanto ad un irrazionale "ritorno al passato"³⁶.

3. Abbiamo davvero bisogno di una responsabilità penale diretta delle IA?

L'idea di concepire una responsabilità penale diretta delle macchine per i danni che possono causare, tuttavia, è probabilmente non solo difficilmente praticabile per le ragioni poc'anzi evidenziate, ma anche fondata su basi poco solide. Questa idea, infatti, sembra partire da una serie di premesse implicite e strettamente intrecciate.

(a) In primo luogo, che un'implementazione delle forme più avanzate di automazione e *machine learning* dovrà necessariamente avvenire, prima o poi, in ogni campo dell'agire umano, a prescindere dal suo "costo" potenziale, a causa della presunta maggiore capacità di queste forme di IA di aumentare i livelli di efficienza, risparmio e sicurezza della società.

(b) Inoltre, che poiché l'"uomo dietro la macchina" non potrebbe esser punito in maniera conforme ai principi fondamentali del diritto penale, quando si utilizza una macchina autonoma e ad autoapprendimento, emergerebbe un *liability gap* a scapito delle vittime³⁷.

(c) Infine, che questo *gap* potrebbe (e dovrebbe) esser necessariamente colmato prevedendo una responsabilità penale diretta a carico delle macchine (e non, ad es., ricorrendo invece ai meccanismi tipici della responsabilità civile).

Purtroppo, le due premesse (a) e (b), su cui si basa l'intero impianto del discorso, che alla fine porta a (c), sembrano fondarsi su un equivoco. La tecnologia non è né buona, né cattiva, né neutrale. Il suo utilizzo non dovrebbe quindi esser ingiustificatamente osteggiato, a causa dell'atavica paura e sfiducia che l'uomo ha sempre nutrito verso le macchine³⁸, ma neppure avventatamente patrocinato, anche quando i rischi derivanti dall'utilizzo di certe IA si rivelano maggiori dei benefici che se ne potrebbero trarre³⁹.

Se partiamo da qui, dall'idea che certi livelli di automazione e l'introduzione di IA intrinsecamente dotate di un pericoloso coefficiente di imprevedibilità dovrebbero ritenersi ammissibili solo nei casi in cui la società possa ritrarre benefici dal loro utilizzo che siano tali da oscurare di gran lunga ogni eventuale pericolosità insita nelle stesse, allora molte questioni diverrebbero probabilmente superflue. Possiamo fare due esempi

³⁵ Cfr. E.P. EVANS, *The Criminal Prosecution and Capital Punishment of Animals*, E.P. Dutton, New York, 1906, *passim*.

³⁶ Possibilista, però, A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, in *Archivio penale*, 3, 2022, p. 17.

³⁷ Cfr. A. MATTHIAS, *The Responsibility Gap. Ascribing Responsibility for the Actions of Learning Automata*, in *Ethics and Information Technology*, 4, 2004, pp. 175 ss.

³⁸ Cfr. N. BOSTROM, *Superintelligenza. Tendenze, pericoli, strategie*, Bollati Boringhieri, Torino, 2023, *passim*; A.K. WOODS, *Robophobia*, in *University of Colorado Law Review*, 93, 2022, pp. 51 ss.

³⁹ Sull'iper-ottimismo di alcuni profeti della "singolarità", cfr. M.A. BODEN, *L'Intelligenza Artificiale*, Il Mulino, Bologna, 2019, pp. 144 ss.

in proposito: quello delle auto a guida autonoma e quello dei sistemi d'arma autonomi (AWS).

Nel primo caso, è generalmente riconosciuto che l'uso di questi veicoli potrebbe in futuro portare ad una drastica riduzione dei rischi di morte e lesioni dovuti alla circolazione stradale⁴⁰. Se ciò è vero (assunto che non dovrebbe esser dato per scontato, ma dimostrato, prove alla mano), dovremmo quindi proibire l'implementazione di tali sistemi, con la conseguente rinuncia ai benefici sociali che essi potrebbero apportare, solo perché possono talvolta condurre anche a taluni esiti imprevedibili? Oppure, dovremmo attribuire la colpa di ogni evento sfortunato e imprevedibile da essi determinato ad attori umani, soltanto per soddisfare gli istinti punitivi e vendicativi della collettività? Queste idee non sembrano avere molto senso. È ovvio che, se i benefici fossero così tanti e così rilevanti, probabilmente si dovrebbe arrivare ad identificare un'area di "rischio lecito" – attraverso un complesso bilanciamento dei benefici collettivi e dei rischi imponderabili propri dei vari sistemi – all'interno della quale al massimo la responsabilità civile dovrebbe esser chiamata a rispondere dei potenziali eventi sfortunati occorsi.

In un futuro in cui la circolazione stradale sarà gestita interamente da sistemi senza conducente, se questo condurrà davvero a una drastica riduzione del numero di incidenti dalle attuali diverse migliaia all'anno ad una manciata di casi (come molti sembrano credere), dovrebbe esser accuratamente scartata l'idea che il rischio della guida completamente autonoma non dovrebbe esser ritenuto ammissibile, o che gli incidenti andrebbero necessariamente addossati penalmente a qualcuno, ritenendoli sempre prevedibili e, come tali, imputabili alla negligenza di qualche soggetto.

In altri campi, tuttavia, si dovrebbe ovviamente giungere a conclusioni diverse. L'uso degli AWS, ad es., sembra sconsigliabile⁴¹ per ragioni deontologiche e morali, ma anche da un punto di vista consequenzialista (cioè basandosi su un mero calcolo costi/benefici). Solo facendo leva su una visione consequenzialista "miope", invero, si potrebbe arrivare a patrocinarne seriamente l'impiego.

Le ragioni consequenzialiste addotte a favore del dispiegamento di AWS, del resto, sono state tipicamente fondate soltanto sull'aspettativa di AWS conformi al diritto internazionale umanitario, in grado di ridurre il numero di vittime tra i combattenti, le vittime innocenti e i danni collaterali, grazie alle loro migliori capacità di *targeting* e di ingaggio, superiori a quelle dei soldati umani, emotivamente fragili e cognitivamente più limitati⁴². Tuttavia, si possono ravvisare ragioni molto più stringenti di questa in favore di un divieto precauzionale di siffatti strumenti tecnologici.

⁴⁰ M. LANZI, *Self-driving cars*, cit., pp. 25 ss.; U. RUFFOLO, *Intelligenza artificiale ed automotive: la responsabilità da veicoli self-driving e driverless*, in U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, p. 155; H. GREGER – J. ESPINO – A.S. SANCHEZ, *Autonomous vehicle: heaven or hell? Creating a transportation revolution that benefits all*, in www.greenlining.org, 2019, *passim*; S. VAN UTYSEL – D. VASCONCELLOS VARGAS, *Challenges for and with Autonomous Vehicles*, in S. VAN UTYSEL – D. VASCONCELLOS VARGAS (eds), *Autonomous Vehicles. Business, Technology and Law*, Springer, Berlin-New York, 2020, 2; J.B. WOODS, *Autonomous Vehicles and Police de-escalation*, in *Northwest Uni Law Review*, 2019, p. 114.

⁴¹ Cfr. N. SHARKLEY, *The Evitability of Autonomous Robot Warfare*, in *International Review of the Red Cross*, 94 (886), 2012, pp. 787 ss.; R. SPARROW, *Killer Robots*, in: *Journal of Applied Philosophy*, 24(1), 2007, pp. 62 ss.

⁴² Cfr. R.C. ARKIN, *Governing Lethal Behavior in Autonomous Robots*, CRC Press, Boca Raton, 2009, *passim*; ID.,

A prescindere da quanto gli AWS riusciranno a svolgere bene i loro compiti di puntamento e ingaggio⁴³, invero, ci sono innanzitutto insuperabili ragioni morali che suggeriscono di proibirne l'uso: gli esseri umani hanno il diritto di non venir privati della propria vita in modo arbitrario, cioè senza il rispetto che gli altri esseri umani devono loro in quanto potenziali vittime della forza letale⁴⁴; quindi, affinché le decisioni di uccidere possano esser considerate non arbitrarie, come sostiene Asaro, dovrebbero sempre venir prese sulla base di un esercizio responsabile del giudizio umano e della compassione⁴⁵.

Poi, è anche certo che un bilanciamento di più ampio respiro dovrebbe indubbiamente portare a riconoscere anche che i rischi associati ad un loro impiego sono tali da superare di gran lunga i benefici che potremmo attenderci da esso⁴⁶. Le ragioni consequenzialiste addotte a sostegno dell'impiego degli AWS, infatti, riguardano solo le prestazioni attese sul campo di battaglia e alcuni dei loro esiti. Vengono però totalmente trascurate le conseguenze che si possono prevedere su scala più globale, come ad es. il fatto che: (a) questo dispiegamento potrebbe avere effetti negativi sulla stabilità della pace; (b) potrebbe incentivare nuove guerre e incoraggiare una *escalation* dalla guerra convenzionale a quella nucleare; (c) e potrebbe, infine, persino disinnescare gli attuali fattori di deterrenza nucleare⁴⁷.

Quindi, per quanto difficile possa rivelarsi il percorso che dovrebbe condurre ad un generalizzato divieto di questi sistemi tecnologici⁴⁸, si tratta comunque, senz'altro, della strada più corretta da percorrere.

Lethal autonomous systems and the plight of the non-combatant, in *AISB Quarterly*, 137, 2013, pp. 1 ss.

⁴³ Ad oggi ci sono fondati dubbi circa il fatto che le stesse possano garantire una completa sicurezza e, al contrario, vi sono prove evidenti di quella che è stata giustamente definita "illusione di precisione": cfr. D. BIRNBACHER, *Are autonomous weapons systems a threat to human dignity?* in N. BUTHA – S. BECK – R. GEIB – H.-Y. LIU – C. KREß (eds), *Autonomous Weapons Systems. Law, Ethics, Policy*, Cambridge University Press, Cambridge, 2016, pp. 117 ss.

⁴⁴ Cfr. C. BAGNOLI, *Teoria della responsabilità*, Il Mulino, Bologna, 2019, p. 78.

⁴⁵ Cfr. P. ASARO, *On banning autonomous weapon systems: human rights, automation and the dehumanisation of lethal decision-making*, in *International Review of the Red Cross*, 94, pp. 687 ss.; D. BIRNBACHER, *Are autonomous weapons systems*, cit., pp. 113 ss.

⁴⁶ Così anche G. TAMBURRINI, *On banning autonomous weapons systems: from deontological to wide consequentialist reasons*, in N. BUTHA – S. BECK – R. GEIB – H.-Y. LIU – C. KREß (eds), *Autonomous Weapons*, cit., pp. 138 ss.

⁴⁷ Cfr. N. SHARKLEY, *Cassandra or the false prophet of doom: Al robots and war*, in *IEEE Intelligent Systems*, 28(4), 2008, p. 16; F. SAUER – N. SCHÖRNIG, *Killer drones: the silver bullet of democratic warfare?*, in *Security Dialogue*, 34, 2012, pp. 363 ss.; J. ALTMANN, *Preventive arms control for uninhabited military vehicles*, in R. CAPURRO – M. NAGENBORG (eds), *Ethics and Robotics*, IOS Press, Amsterdam, 2009, pp. 80 ss.; G. TAMBURRINI, *On banning autonomous weapons systems*, cit., pp. 139 ss.

⁴⁸ Ad es., N. BUTHA – S. BECK – R. GEIB, *Present futures: concluding reflections and open questions on autonomous weapons systems*, in N. BUTHA – S. BECK – R. GEIB – H.-Y. LIU – C. KREß (eds), *Autonomous Weapons*, cit., p. 378 ritengono che, «nella fase attuale, sembra improbabile e addirittura irrealistico che un divieto più pervasivo di tecnologie autonome nei sistemi militari possa vedere la luce».

4. Alcune osservazioni conclusive.

Sembra dunque evidente che, invece di continuare ancora a lungo a ipotizzare una responsabilità penale diretta delle macchine, ci si dovrebbe forse limitare a fare una selezione più attenta delle varie attività umane in cui l'impiego di sistemi altamente automatizzati potrebbe ritenersi veramente "consigliabile"⁴⁹. Eliminare dall'equazione l'idea che si possa (o persino si debba) concepire una responsabilità penale diretta delle IA, infatti, permetterebbe di concentrare tutti gli sforzi e le attenzioni sulle questioni problematiche che sembrano contare davvero per il diritto penale, dirottando così tutte le energie verso la loro risoluzione.

In fin dei conti, l'uso di IA completamente autonome e probabilmente a volte anche la ricerca scientifica sul loro sviluppo dovrebbero ammettersi solo quando si può dimostrare che la società – alla luce di un'adeguata analisi costi/benefici – può trarre più vantaggi che svantaggi dal loro impiego. Nelle aree in cui l'uso di queste IA è ritenuto troppo rischioso, e non così benefico o in grado di garantire il dovuto rispetto di alcune prerogative umane fondamentali, i precetti penali dovrebbero, semmai, intervenire al fine: (a) o di regolamentare e all'occorrenza vietare (sul modello dell'*Embryo Protection Act*) la ricerca e/o l'introduzione nel sistema di questi strumenti "pericolosi"⁵⁰; (b) o di stabilire, comunque, che il mantenimento di un controllo umano significativo debba ritenersi essenziale in materia, garantendo così un centro d'imputazione umano a cui eventualmente addossare la responsabilità.

D'altra parte, nelle aree in cui tali tecnologie altamente autonome sono ritenute accettabili e i rischi tutto sommato tollerabili (nella misura in cui sono controbilanciati dai grandi benefici che si possono ricavare dal loro utilizzo), si dovrebbe verosimilmente riconoscere che all'interno dell'area di "rischio socialmente consentito" delineata da ciascun sistema⁵¹, di norma non dovrebbe sorgere alcun tipo di responsabilità penale per i programmatori, i produttori o gli utenti, e che gli eventuali (rari) eventi sfortunati occorsi dovrebbero verosimilmente restare vieppiù appannaggio dei meccanismi tipici della responsabilità civile⁵².

⁴⁹ Ad es. si potrebbe verificare se tale sia il caso anche per il settore medico o per quello dei veicoli marittimi: U. PAGALLO, *Il dovere alla salute. Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario*, Mimesis, Milano, 2022, *passim*; N. KLEIN – D. GUILFOYLE – M.S. KARIM – R. MCLAGHLIN, *Maritime Autonomous Vehicles: New Frontiers in the Law of the Sea*, in *International and Comparative Law Quarterly*, 69(3), 2000, pp. 719 ss.

⁵⁰ Cfr. M.B. MAGRO, *Decisione umana*, cit., pp. 20 ss.; K. GAEDE, *Künstliche Intelligenz - Rechte und Strafen für Roboter? Plädoyer für eine Regulierung künstlicher Intelligenz jenseits ihrer reinen Anwendung*, Nomos, Baden-Baden, 2019, pp. 76 ss. e pp. 81 ss.; G. QUINTERO OLIVARES, *La robótica*, cit., pp. 21-22.

⁵¹ Ma sarebbe ancora meglio se quest'area di rischio venisse delineata a livello globale nel modo più uniforme possibile, al fine di evitare le discrepanze nella classificazione dei casi *cross-border* che, come osservato da V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. RUFFOLO (a cura di), *Intelligenza artificiale*, cit., pp. 548 ss., potrebbero porsi.

⁵² Cfr. 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 (a cura di), *Il diritto nell'era digitale*, cit., pp. 719 ss.; R. BORGOGNO, *La responsabilità penale nei processi ad elevata automazione*, *ivi*, 735 ss.; A. FIORELLA, *Responsabilità penale del Tutor*, cit., pp. 654 ss.; R. BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *www.mediaLaws.eu*, 20.11.2019, p. 268.

Quel che è certo, poi, è che il potenziale intervento del diritto penale in materia (laddove ritenuto essenziale) dovrebbe comunque restare permanentemente incentrato su responsabilità umane (al più chiamando in causa anche la responsabilità da reato degli enti)⁵³, orientandosi principalmente verso la punizione di violazioni di regole di condotta stabilite *ex ante*, volte a prevenire, anticipare e minimizzare i profili di rischio che la tecnologia più autonoma può presentare. Come si è osservato, infatti, «se le categorie penalistiche non possono essere slabbrate per imputare eventi imponderabili, ci si deve limitare ad aggredire, con pene proporzionate», il «disvalore di condotte» che possono costituire forme «di *miscompliance* individuale rispetto a protocolli preventivi nell'impiego di strumenti tecnologicamente avanzati»⁵⁴. Al diritto penale non si potrà chiedere molto più di questo.

⁵³ Cfr. A.F. TRIPODI, *Uomo, Societas, Machina*, in www.la-legislazionepenale.eu, 10.5.2023, pp. 12 ss.

⁵⁴ Cfr. F. CONSULICH, *Flash Offenders*, cit., pp. 1053 ss.

Editore

ASSOCIAZIONE
**"PROGETTO GIUSTIZIA
PENALE"**