

## L'INTELLIGENZA ARTIFICIALE: UN BANCO DI PROVA DELLE CATEGORIE TRADIZIONALI DELLA RESPONSABILITÀ PENALE

Isabella Fontana \*

**SOMMARIO:** 1.- Introduzione; 2.- Il diritto penale alla prova dell'intelligenza artificiale: definizioni e scenari; 3.- L'IA come strumento delinquente e la responsabilità dolosa; 4.- L'IA come fonte di rischio tecnologico e la responsabilità colposa; 5.- Come sfuggire ai vuoti di responsabilità penale: le soluzioni della dottrina; 6.- Osservazioni conclusive.

### 1.- Introduzione.

L'intelligenza artificiale (IA) rappresenta un'innovazione tecnologica senza precedenti, destinata ad incidere in modo significativo su ogni ramo del diritto, a partire da quella branca deputata, tradizionalmente, a relazionarsi con gli esseri umani: il diritto penale. A ben guardare, il sistema penale, dalla previsione astratta del reato fino alla concreta irrogazione di una pena, all'esito di un processo regolato per legge, risulta edificato proprio sull'essere umano<sup>1</sup>. Nucleo essenziale di ogni fattispecie incriminatrice è l'azione dell'uomo - intesa come volontà libera che si realizza nel mondo esterno - oggetto di un rimprovero personale e colpevole (artt. 25, comma 2, e 27, comma 1, Cost.; art. 7 CEDU; art. 49 CDFUE). Ancora, il processo penale, la sede istituzionalmente deputata all'accertamento giudiziale dei fatti di reato (art. 111 Cost.; art. 6 CEDU; artt. 47 e 48 CDFUE), si risolve in un giudizio affidato ad un uomo su un altro uomo, quale esperienza unica e irripetibile. Ed infine, la sanzione punitiva, sia pure informata al principio del finalismo rieducativo (art. 27, comma 3, Cost.), rimane una afflizione ineliminabile per il suo naturale destinatario umano<sup>2</sup>. Questo, in breve, il diritto penale che viene restituito dalla Carta costituzionale e dai principali testi internazionali in materia di diritti fondamentali. Un sistema "personalistico e personologico"<sup>3</sup>, dunque, già abituato a misurarsi con le sfide poste dall'evoluzione tecnologica, tipiche della moderna "società del rischio"<sup>4</sup>, riuscendo a mantenere salde le sue fondamenta<sup>5</sup>.

Tuttavia, è opinione diffusa che lo sviluppo dell'IA e il suo crescente impiego in moltissimi settori - come la circolazione stradale<sup>6</sup>, la sanità<sup>7</sup> e il mercato finanziario<sup>8</sup> - conducano ad un rapporto uomo-

\* Dottoressa di ricerca in diritto penale presso l'Università degli Studi "Roma Tre".

<sup>1</sup> Così V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen* (2020) 2ss. Cfr. M. Lanzi, *Uso di strumenti a Intelligenza Artificiale e imputazione della responsabilità penale: difficoltà ricostruttive e prospettive di intervento*, in *Diritto di Difesa* 3-4 (2023) 775ss.

<sup>2</sup> In questi termini M. Trapani, *Il reato e le sue conseguenze. Punibilità, pena, punizione in un sistema criminale integrale e integrato*, Roma 2022, 54.

<sup>3</sup> V. Manes, *L'oracolo* cit. 2.

<sup>4</sup> U. Beck, *La società del rischio. Verso una seconda modernità*, Roma 2017.

<sup>5</sup> In argomento, F. Stella, *Giustizia e modernità. La protezione dell'innocente e tutela delle vittime*, Milano 2003, 3ss.; G. Marinucci, *Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.* (2005) 29ss.

<sup>6</sup> Nello specifico, si rinvia a M. Lanzi, *Self-driving cars e responsabilità penale. La gestione del "rischio stradale" nell'era dell'intelligenza artificiale*, Torino 2023, 23ss.

<sup>7</sup> Sui rapporti tra l'IA e l'attività sanitaria si veda R. Scotti, *La responsabilità civile dei danni cagionati da sistemi di intelligenza artificiale in ambito sanitario*, in *Giustizia civile* 1 (2024) 158ss.

<sup>8</sup> Per tutti, F. Consulich, *Il nastro di Möbius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso di mercato*, in *Banca Borsa Titoli di Credito* 2 (2018) 195ss.

macchina inedito, che non può essere letto e interpretato secondo gli schemi tradizionali del passato<sup>9</sup>. Se, da un lato, il diritto penale non può rinunciare alla sua funzione essenziale di tutela *in extremis* di beni fondamentali contro le aggressioni illecite da chiunque commesse - inclusi gli algoritmi intelligenti - dall'altro, non si può negare che le proprietà, senza dubbio, uniche dei sistemi di IA mettano in discussione le categorie dogmatiche tradizionali.

Il presente contributo si propone, quindi, di esaminare i profili di interferenza tra il diritto penale e l'intelligenza artificiale, con l'obiettivo di valorizzare le problematiche in punto di imputazione della responsabilità penale e le possibili soluzioni escogitate dalla scienza penalistica, nazionale e internazionale, per scongiurare vuoti di tutela socialmente inaccettabili.

## **2.- Il diritto penale alla prova dell'intelligenza artificiale: definizioni e scenari.**

Per un corretto approccio metodologico al tema, appare indispensabile muovere da una premessa di tipo definitorio: cosa si intende per intelligenza artificiale.

Nella letteratura scientifica di riferimento, manca una definizione universalmente accettata di IA<sup>10</sup>. Ai fini della presente indagine, l'intelligenza artificiale può essere definita come la tecnologia che consente alle macchine e ai computer di simulare attività tipicamente umane, tra cui l'apprendimento, il "decision making", la creatività e l'autonomia<sup>11</sup>.

Ad oggi, i sistemi di IA sono capaci di memorizzare ed elaborare enormi quantità di dati<sup>12</sup> - forniti dall'operatore o raccolti dalla macchina - da cui inferiscono le informazioni per formulare previsioni, prendere decisioni e intraprendere azioni nell'ambiente reale o virtuale di inserimento, così da risolvere in modo ottimale il compito assegnato in sede di programmazione<sup>13</sup>. A seconda della capacità di interazione con l'uomo, i sistemi di IA si classificano in due tipologie: "deboli" e "forti". I primi si basano sull'esecuzione di programmi predefiniti, destinati a rielaborare un quantitativo predeterminato di dati e informazioni immesse dagli operatori<sup>14</sup>. I secondi, invece, sono dotati di capacità di apprendimento automatico (il "machine learning")<sup>15</sup> e sono in grado di adattarsi ed

<sup>9</sup> Così C. Burchard, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.* 4 (2019) 1908ss.; G. F. Italiano, *Intelligenza artificiale, che errore lasciarla agli informatici*, in *Agenda Digitale* (2019) 7ss.

<sup>10</sup> Cfr. Lanzi, *Self-driving cars* cit. 13ss.; B. Panattoni. *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Diritto dell'Informazione e dell'Informatica* 2 (2021) 317ss. Una definizione normativa è contenuta all'art. 3, lett. 1), del Regolamento europeo 1689 del 13/06/2024, dove per sistema di IA si intende «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi esplicativi o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

<sup>11</sup> C. Stryker, *What is artificial intelligence (AI)*, in [www.ibm.com](http://www.ibm.com). Il termine "intelligenza artificiale" fu coniato, per la prima volta, nel 1956, dal matematico americano John McCarthy, in occasione della prima conferenza tenutasi negli Stati Uniti, al Dartmouth College del New Hampshire. McCarthy, ad oggi considerato il padre della materia, così definisce l'intelligenza artificiale "it is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable". J. McCharty, *What is artificial intelligence*, Standford University 2004, 2ss.

<sup>12</sup> Il carburante necessario per allenare gli attuali sistemi di IA (il c.d. "training") è costituito proprio dai "Big data". Sul tema A. Longo, G. Scorsa, *Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà*, Milano 2020, 8ss.

<sup>13</sup> A tal proposito, si distingue tra l'IA ristretta, che comprende i sistemi progettati e utilizzati per svolgere compiti specifici, anche se complessi, come guidare una macchina o eseguire transazioni finanziarie, e l'IA generale, che indica un sistema capace di risolvere qualsiasi compito assegnato. Così S. Quintarelli, *Intelligenza Artificiale. Cos'è davvero, come funziona, che effetti avrà*, Torino 2020, 36ss.

<sup>14</sup> Sul tema, F. Consulich, M. Maugeri, C. Milia, T.N. Poli, G. Trovatore, *AI e abusi di mercato: le leggi della robotica si applicano alle operazioni finanziarie*, in *Quaderni giuridici Consob* 29 (2023) 23ss.; Lanzi, *Self-driving* cit., 14ss.

<sup>15</sup> Sulle tecniche di "machine learning" si rinvia a H. Surden, *Machine learning and Law*, in *Washington Law Review* 87(2014) 88ss.

evolversi in base all'esperienza acquisita, generando risultati imprevedibili e autonomi rispetto agli "inputs" iniziali. Non solo. A differenza degli algoritmi tradizionali, questo secondo tipo di IA segue meccanismi decisionali di per sé poco trasparenti e non immediatamente decifrabili, funzionando come una sorta di scatola chiusa (la "black box")<sup>16</sup>.

Senza dubbio, sono i sistemi di IA "forti" ad assumere particolare interesse per il penalista, in ragione delle richiamate caratteristiche strutturali, come l'autoapprendimento e l'opacità algoritmica, che li rendono artefatti tecnologici imprevedibili e imponderabili "by design"<sup>17</sup>.

Si tratta di esaminare, ora, a che titolo questi sistemi possono venire coinvolti nella commissione di fatti dotati di rilevanza penale. In via esemplificativa, possono distinguersi due scenari<sup>18</sup>:

- a) il sistema di IA è lo strumento nelle mani dell'essere umano, tramite il quale viene commesso il reato (ad es. un algoritmo viene istruito per commettere manipolazioni di mercato o frodi informatiche);
- b) il sistema di IA è l'autore principale del fatto costitutivo dell'offesa incriminata (ad es. una macchina a guida autonoma investe una persona cagionandone la morte).

### **3.- L'IA come strumento delinquente e la responsabilità dolosa.**

Il primo scenario riguarda, essenzialmente, i reati a matrice dolosa commessi dall'uomo tramite i sistemi di IA<sup>19</sup>. Qui, non si pongono particolari problemi a riconoscere responsabile a titolo di dolo la persona che "sta dietro" la macchina. È pacifico, infatti, che il reato sia direttamente imputabile all'uomo quando l'intelligenza artificiale è utilizzata come *instrumentum sceleris* per conseguire il risultato criminoso voluto dall'agente (il c.d. algoritmo delinquente)<sup>20</sup>. Entro questa prospettiva, si ritiene che le proprietà che qualificano i moderni sistemi di IA non rappresentino un reale ostacolo alla operatività dei meccanismi ascrittivi della responsabilità dolosa, comprese le regole che disciplinano le ipotesi di reato aberrante (artt. 82 e 83 c.p.)<sup>21</sup>. A fronte di un "input" umano diretto alla commissione di un reato, eventuali deviazioni nel concreto decorso causale riconducibili al comportamento imprevedibile dell'algoritmo non spezzerebbero né il collegamento attuale con l'offesa né la sua riferibilità soggettiva all'uomo, risolvendosi, al più, in una *aberratio causae* irrilevante<sup>22</sup>.

Eppure, anche in casi semplici, come questi, l'impiego di algoritmi intelligenti potrebbe sollevare significative problematiche sulla prova dell'elemento soggettivo in capo all'agente umano<sup>23</sup>. Nel giudizio penale, come noto, perché si raggiunga la prova del dolo di un determinato reato, non basta dimostrare una generica e astratta intenzione delittuosa, ma l'effettiva volizione e rappresentazione dell'evento dannoso, come concretamente verificatosi *hic et nunc*<sup>24</sup>. Ecco allora che, quando la

<sup>16</sup> Italiano, *Intelligenza artificiale* cit. 5ss.

<sup>17</sup> R. Calo, *Robotics and the Lessons of Cyberlaw*, in *103 California Law Review* (2015) 542.

<sup>18</sup> In argomento, cfr. F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo* (2019) 24ss.; Lanzi, *Uso di strumenti* cit. 781ss.

<sup>19</sup> Così A. Cappellini, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Discrimen* (2019) 7ss.

<sup>20</sup> L. Picotti, *Categorie tradizionali del diritto penale e intelligenza artificiale: crisi o palingenesi? La raccomandazione dell'Association Internationale de Droit Pénal e la rilevanza del recente regolamento europeo sull'intelligenza artificiale*, in *Sistema penale* (2024) 10ss.

<sup>21</sup> Cappellini, *Machina* cit. 8.

<sup>22</sup> Per tutti A. Pagliaro, *Il reato. Parte generale*, Milano 2007, 442ss.

<sup>23</sup> Così F. Consulich, *Le prospettive di Accountability penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.* (2022) 1031ss.; Id, *Il nastro* cit. 218ss.

<sup>24</sup> L. Romanò, *La responsabilità penale al tempo di Chatgpt: prospettive de iure condendo in tema di gestione del rischio da intelligenza artificiale generativa*, in *Sistema penale* (2023) 12ss.

consumazione del reato è rimessa alla discrezionalità operativa di una macchina autonoma, che decide quando, come e con quale intensità agire, potrebbe mancare una piena consapevolezza dell’evoluzione della condotta algoritmica in capo al singolo individuo, con conseguente esclusione di responsabilità per difetto di colpevolezza<sup>25</sup>.

Un ulteriore profilo di interferenza riguarda le ipotesi in cui l’intelligenza artificiale costituisce, per sé stessa, il prodotto di una attività umana vietata dalla legge (ad es. un algoritmo viene addestrato con dati acquisiti in modo illecito oppure viene programmato per violare sistemi informatici protetti)<sup>26</sup>. Qui, assume rilievo centrale l’attività, esclusivamente umana, del programmare, sviluppare o utilizzare coscientemente algoritmi intelligenti che, in ragione del loro funzionamento o del compito assegnato, concretizzano un’offesa a beni giuridici già tutelati dall’ordinamento - come la privacy e l’integrità dei sistemi informatici<sup>27</sup> - oppure in via di emersione, come l’affidabilità dei sistemi di IA<sup>28</sup>.

#### **4.- L’IA come fonte di rischio tecnologico e la responsabilità colposa.**

Il secondo scenario interessa le ipotesi in cui il sistema di IA, progettato per finalità lecite o impiegato in contesti autorizzati, cagiona un’offesa per un percorso decisionale autonomo innescato dalla macchina, terreno di elezione delle forme di imputazione colposa. Sotto questo profilo, i sistemi di IA vengono in rilievo come fonte di rischio per i beni giuridici protetti dal diritto penale, come la vita o l’incolumità individuale, che devono essere governati e gestiti dagli operatori al fine di prevenire e minimizzare i danni che ne possono derivare.

Sennonché, gli algoritmi intelligenti, a differenza di altri artefatti tecnologici, presentano caratteristiche che compromettono l’applicazione dei tradizionali meccanismi ascrittivi della imputazione colposa, generando veri e propri vuoti di responsabilità (c.d. “responsibility gap”)<sup>29</sup>. Entrando nel merito, la scienza penalistica segnala diversi momenti di frizione con lo statuto penale della colpa. In primo luogo, viene evidenziata la difficoltà di individuare un centro di imputazione umano responsabile per il singolo evento dannoso: questi dispositivi, infatti, sono programmati, sviluppati e messi in commercio da una filiera produttiva complessa, in cui intervengono una pluralità di individui e organizzazioni, anche di grandi dimensioni, con conseguente frantumazione dei rispettivi ruoli<sup>30</sup>. In secondo luogo, il settore dell’IA è sottoposto a continui e rapidi sviluppi, il che renderebbe quasi impossibile per il legislatore “fissare” un catalogo normativo di regole cautelari che siano al passo con l’evoluzione tecnologica<sup>31</sup>. Ancora, l’imprevedibilità strutturale del

<sup>25</sup> Cfr. Id, *La responsabilità* cit. 14; Consulich, *Il nastro* cit. 219.

<sup>26</sup> P. Troncone, *Il sistema dell’intelligenza artificiale nella trama grammaticale del diritto penale. Dalla responsabilità umana alla responsabilità delle macchine pensanti: un inatteso return trip effect*, in *Cass. pen.* 9 (2022) 3287ss.

<sup>27</sup> Così F. Lagioia, G. Sartor, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *Philosophy and Technology* (2020) 456ss.

<sup>28</sup> B. Panattoni, *Generative AI and criminal law*, in *Cambridge Forum on AI: Law and Governance* (2025) 5.

<sup>29</sup> Esulano dalla problematica in commento i casi in cui sia possibile stabilire un collegamento eziologico certo tra l’evento dannoso e l’inoservanza di una regola cautelare, scritta o non scritta, da parte dell’operatore umano. Sotto questo profilo, vengono in rilievo le ipotesi classiche di colpa del produttore, quando il danno dipende da un difetto di produzione, di manutenzione oppure da un errore di programmazione dell’algoritmo; nonché, le ipotesi di colpa dell’utilizzatore, quando l’evento lesivo deriva da un uso incauto della macchina o dalla mancata attivazione delle misure di sicurezza. Per tutti A. Cappellini, *Reati colposi e tecnologie dell’intelligenza artificiale*, in G. De Francesco (cur.), *Il diritto penale di fronte alle sfide della «società del rischio»: un difficile rapporto tra nuove esigenze di tutela e classici equilibri di sistema*, Torino 2018, 23ss.

<sup>30</sup> Cfr. Cappellini, *Machina* cit. 26; I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.* 1 (2021) 83ss.

<sup>31</sup> Lanzi, *Uso di strumenti* cit. 784ss.

comportamento algoritmico appare inconciliabile con il giudizio di prevedibilità ed evitabilità, *ex ante* ed in concreto, dell'evento dannoso, necessario per integrare il momento soggettivo della colpa<sup>32</sup>.

Problematiche simili si incontrano anche sul terreno della responsabilità penale di tipo omissivo. In linea di principio, si osserva, i controllori umani non sarebbero in grado di percepire l'anomalia o il fallimento del sistema di IA, a causa della opacità algoritmica e della complessità tecnica della materia, né, tantomeno, di intervenire “in corso d’opera” al fine di neutralizzare o correggere macchine progettate per sostituirli<sup>33</sup>. A ben guardare, si tratta di ostacoli difficilmente superabili, se non a patto di forzare i limiti delle categorie tradizionali e trasfigurare la colpa penale in una inammissibile forma di responsabilità oggettiva, con il solo scopo di offrire alla collettività un capro espiatorio in carne ed ossa<sup>34</sup>.

### **5.- Come sfuggire ai vuoti di responsabilità penale: le soluzioni della dottrina.**

Nella prospettiva di colmare i vuoti di responsabilità penale che si potrebbero generare a fronte di eventi lesivi algoritmici<sup>35</sup>, la scienza penalistica, sia nazionale che internazionale, ha escogitato alcune soluzioni per orientare il legislatore del prossimo futuro.

Una prima teoria, di origine americana, propone di introdurre una responsabilità diretta delle macchine, da riconoscere, al pari delle persone giuridiche, come soggetti destinatari di obblighi e divieti penalmente sanzionati<sup>36</sup>. Secondo questo modello, i sistemi di IA più avanzati, dotati di capacità cognitive e decisionali autonome, sarebbero in grado di soddisfare entrambi i requisiti, oggettivi e soggettivi, necessari per imputare un fatto di reato (rispettivamente, l'*actus reus* e la *mens rea*)<sup>37</sup>. A ciò si aggiunge che le diverse tipologie di sanzioni punitive, sia personali che reali, previste negli ordinamenti moderni, ben si potrebbero applicare nei confronti delle macchine, con i dovuti adeguamenti<sup>38</sup>. Contro questa tesi, rimasta minoritaria nel panorama scientifico europeo<sup>39</sup>, si sollevano, principalmente, tre ordini di obiezioni: l’assenza di rimproverabilità e di colpevolezza in capo ai sistemi di IA, la totale perdita di senso della sanzione punitiva nei confronti delle macchine e

---

<sup>32</sup> Cfr. Cappellini, *Reati colposi* cit. 25ss.

<sup>33</sup> Così Romanò, *La responsabilità penale* cit. 16ss. In dottrina, oltre all’opacità algoritmica, si segnalano i seguenti fattori problematici, connessi all’autonomia dei sistemi di IA, che potrebbero precludere una imputazione a titolo omissivo nei confronti del sorvegliante umano: la c.d. distorsione dell’automazione, la tendenza, cioè, delle persone ad affidarsi eccessivamente alla tecnologia e a non metterne in discussione i risultati; la ridotta capacità di reazione nelle situazioni di emergenza, dovuta al fatto che il livello di attenzione di chi si limita a monitorare l’attività altrui è inferiore rispetto a quello di chi agisce in prima persona (il c.d. “control dilemma”). In argomento cfr. Panattoni, *Intelligenza artificiale* cit. 337; A. Giannini, *Intelligenza artificiale, human oversight e responsabilità penale: prove d’impatto a livello europeo*, in *Discrimen* (2022) 23.

<sup>34</sup> Cappellini, *Reati colposi* cit. 31.

<sup>35</sup> Attualmente, uno dei principali settori colpiti dai danni algoritmici è quello delle transazioni finanziarie. Sul punto, F. Consulich, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.* 3 (2022) 1015ss.

<sup>36</sup> Si tratta della tesi sostenuta da G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities*, in *Social Science Research Network* (2010) 9ss.

<sup>37</sup> Id, *The Criminal* cit. 22ss.

<sup>38</sup> Id, *The Criminal* cit. 33ss., secondo il quale lo stesso risultato della pena detentiva potrebbe essere raggiunto nei confronti delle macchine tramite una sanzione che vietи il loro utilizzo per un determinato periodo di tempo; per la pecuniaria, invece, sarebbe sufficiente dotare i sistemi di IA di un patrimonio autonomo.

<sup>39</sup> Picotti, *Categorie tradizionali* cit. 7.

l'infondatezza teorica del parallelismo con le persone giuridiche, pur sempre costituite da esseri umani<sup>40</sup>.

Fermo, dunque, il principio assiologico per cui *machina delinquere non potest*<sup>41</sup>, un secondo indirizzo propone di implementare il modello della responsabilità da reato dell'ente, per superare le difficoltà legate all'attribuzione di un coefficiente di colpevolezza in capo ai singoli individui<sup>42</sup>. Anche in mancanza di una persona fisica responsabile per l'imprevedibilità della deviazione algoritmica, si potrebbe formulare un rimprovero a titolo di "colpa di organizzazione" nei confronti dell'ente tutte le volte in cui vi sia stato un impiego di agenti artificiali malevoli o eccessivamente rischiosi in assenza di predeterminate cautele doverose<sup>43</sup>.

Per una terza linea interpretativa, dotata di maggiori consensi in dottrina, il modello classico del "diritto penale dell'evento" sarebbe del tutto inadeguato per dare copertura agli eventi lesivi cagionati dai sistemi di IA più sofisticati. Di qui, la necessità di virare verso forme di anticipazione della responsabilità penale, riconducibili al paradigma opposto del "diritto penale del comportamento"<sup>44</sup>. Entro questa prospettiva, si ritiene che una risposta sanzionatoria efficace possa derivare - oltre che dal ricorso ai reati di pericolo astratto - dal noto modello ingiunzionale<sup>45</sup>. In breve, le pubbliche autorità intimano agli operatori del settore di adottare determinate misure di sicurezza (ad es. sottoporre gli algoritmi ad una sperimentazione aggiuntiva, disattivarli oppure rimuoverli dal mercato); in caso di inosservanza dolosa o colposa di tali prescrizioni, scattano le sanzioni punitive<sup>46</sup>. Sarà, poi, compito del diritto civile e amministrativo fornire tutela risarcitoria alle vittime degli eventi avversi derivanti dalle deviazioni imprevedibili degli algoritmi, così da salvaguardare, ad un tempo, il carattere residuale del diritto penale e l'integrità delle categorie dogmatiche.

## 6.- Osservazioni conclusive.

In definitiva, tutti concordano sulla necessità di un intervento punitivo di fronte ai rischi tecnologici delle IA, che minacciano in modo durevole i beni giuridici fondamentali dei singoli individui. I tempi e le modalità di questo intervento, nelle moderne democrazie occidentali, sono di prerogativa esclusiva del decisore politico, chiamato a svolgere, nel prossimo futuro, delicate operazioni di bilanciamento tra gli interessi in gioco: da un lato, i benefici apportati dai sistemi di IA, dall'altro i rischi avversi sulla collettività<sup>47</sup>. Un bilanciamento, dunque, tra il principio di precauzione e il principio di prevenzione<sup>48</sup>, da svolgere senza perdere di vista il ruolo costituzionale di *extrema ratio* assegnato al diritto punitivo<sup>49</sup>.

<sup>40</sup> Cfr. Cappellini, *Machina* cit. 10ss.; R. Borsari, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *Medialaws* (2019) 266ss.; C. Piergallini, *Intelligenza artificiale: da "mezzo" ad "autore" del reato*, in *Riv. it. dir. proc. pen.* 4(2020) 1768ss.; L. Greco, *Diritto penal para robôs*, in [www.jota.info](http://www.jota.info).

<sup>41</sup> L'assioma è di Cappellini, *Machina* cit. 19.

<sup>42</sup> Consulich, *Flash offenders* cit. 1040.; Id., *Il nastro* cit. 224ss.

<sup>43</sup> In senso critico, Cappellini, *Reati colposi* cit. 34ss., secondo cui l'imprevedibilità del comportamento algoritmico non consentirebbe di formulare un rimprovero a titolo di colpa neanche nei confronti delle persone giuridiche.

<sup>44</sup> Cfr. Consulich, *Flash offenders* cit. 1042. B. Fragasso, *Intelligenza artificiale e crisi del diritto penale d'evento: profili di responsabilità penale del produttore di sistemi di I.A.*, in *Riv. it. dir. proc. pen.* 1 (2024) 301ss.

<sup>45</sup> Sul modello ingiunzionale si rinvia a Marinucci, *Innovazioni tecnologiche* cit. 57ss.

<sup>46</sup> Piergallini, *Intelligenza artificiale* cit. 1745ss.

<sup>47</sup> F. Basile, B. Fragasso, *Intelligenza artificiale e diritto penale: prove tecniche di convivenza*, in *Diritto di Difesa* (2023) 752ss.

<sup>48</sup> G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Diritti penale contemporaneo* 4 (2020) 80ss.

<sup>49</sup> M.E. Florio, *Il dibattito sulla responsabilità penale diretta delle IA: "molto rumore per nulla"*, in *Sistema penale* 2 (2024) 17ss.; Cfr. Borsari, *Intelligenza Artificiale* cit. 268.

A livello sovranazionale, un primo modello di regolazione di tipo orizzontale e antropocentrico, parametrato sui rischi per i diritti fondamentali delle persone, viene offerto dal recente Regolamento europeo sull'IA. Pur senza contenere esplicativi riferimenti alla materia penale, la normativa europea fornisce ai legislatori nazionali un apparato di obblighi di trasparenza, pubblicità e di gestione del rischio gravanti sui diversi operatori coinvolti, che potrebbero costituire la base per l'introduzione di nuove fattispecie incriminatrici<sup>50</sup>. Nella medesima prospettiva, si inseriscono anche le indicazioni provenienti dal Consiglio d'Europa, dove si raccomanda agli Stati membri di incriminare la progettazione e lo sviluppo dei sistemi di IA in violazione di standard di sicurezza, soprattutto quando l'inosservanza di tali misure comporti la diffusione di informazioni false e fuorvianti, con impatti negativi sui diritti umani, la democrazia e lo Stato di diritto<sup>51</sup>.

La letteratura scientifica, dal canto suo, ha già da tempo evidenziato le criticità sulla tenuta delle categorie penalistiche dinanzi all'impatto irruento dei sistemi di IA nel nostro tessuto sociale. L'auspicio condiviso è che le pubbliche autorità non cedano a un approccio eccessivamente attendista<sup>52</sup> - come quello che ha caratterizzato le piattaforme digitali<sup>53</sup> - e optino per un modello di regolazione proattivo, in cooperazione con i protagonisti del conflitto sociale, a partire dai "Big Tech", depositari di un sapere scientifico e tecnologico di gran lunga superiore a quello degli Stati nazionali<sup>54</sup>.

Preso atto che "la tecnologia non è né buona né cattiva, ma non è neanche neutrale"<sup>55</sup>, appare indispensabile che il rapporto tra persone e macchine si fondi sul rispetto dei diritti fondamentali e che i risvolti normativi delle moderne tecnologie siano valutati secondo i protocolli di un sistema democratico, unico modo, questo, "per sfuggire alle contrapposizioni tra utopie e distopie, tra esaltazione e rifiuto della dimensione tecnologica"<sup>56</sup>.

Anche perché, l'intelligenza artificiale è solo l'inizio<sup>57</sup>.

**Abstract.-** Il contributo esamina come l'impiego di sofisticati sistemi di IA metta in crisi le categorie tradizionali di attribuzione della responsabilità penale, a partire dalla colpevolezza. Le proprietà che caratterizzano gli agenti artificiali c.d. forti (l'autoapprendimento, l'opacità e l'imprevedibilità algoritmica) sollevano numerosi interrogativi sia sul terreno della responsabilità dolosa, quando le IA sono impiegate o progettate per finalità illecite, che di quella colposa, quando l'utilizzo delle macchine rientra nell'area di rischio consentito. La scienza penalistica, a livello nazionale e internazionale, ha elaborato diverse soluzioni per scongiurare vuoti di responsabilità penale a fronte di eventi lesivi algoritmici, arrivando persino a teorizzare una capacità penale diretta dei sistemi di

<sup>50</sup> In una prospettiva *de jure condendo*, parte della dottrina propone di elevare a reati autonomi le condotte di dolosa progettazione, sviluppo e diffusione di sistemi di IA vietati dal Regolamento europeo, perché integrativi di "rischi inaccettabili". Così Picotti, *Categorie tradizionali* cit.; Panattoni, *Generative AI* cit.; Fragasso, *La responsabilità penale* cit. 20ss.

<sup>51</sup> European Committee on Crime Problems, *Discussion Paper on Criminal Liability Related to AI systems - following up Framework Convention on AI & Human Rights, Democracy, the Rule of Law*, in [www.coe.int](http://www.coe.int).

<sup>52</sup> Cappellini, *Reati colposi* cit. 31ss.

<sup>53</sup> Sul tema si rinvia a E. Pedemonte, *La fattoria degli umani. Come le piattaforme digitali stanno riprogettando la nostra vita*, Roma 2024.

<sup>54</sup> Così Piergallini, *Intelligenza artificiale* cit. 1771ss.

<sup>55</sup> M. Kranzberg, *Technology and History: "Kranzberg's Laws"*, in *Technology and Culture* 3 (1986) 544.

<sup>56</sup> Fondamentali, ancora oggi, le riflessioni di S. Rodotà, *Il diritto di avere diritti*, Bari 2012, 340ss.

<sup>57</sup> Così A. Webb, *Why "Living Intelligence" Is The Next Big Thing*, in *Harvard Business Review* (2025) 1ss., la quale osserva come l'intersezione tra l'intelligenza artificiale, i sensori avanzati e le biotecnologie sia destinata a plasmare una nuova realtà, definita come "intelligenza vivente", comprensiva di sistemi capaci, non solo di percepire l'ambiente circostante, ma anche di apprendere, adattarsi ed evolversi in modo simile agli organismi viventi.

IA. Di qui, l'urgenza di un intervento legislativo che sappia bilanciare gli interessi in gioco, salvaguardando il ruolo costituzionale di *extrema ratio* del diritto punitivo.

This contribution examines how the deployment of sophisticated AI systems poses challenges to traditional categories of criminal liability, beginning with the concept of culpability. The defining features of so-called strong artificial agents (such as machine learning, opacity, and algorithmic unpredictability) raise complex issues with respect to both intentional liability, when AI is designed or employed for unlawful purposes, and negligent liability, when its use falls within the bounds of permissible risk. Criminal law scholarship, both nationally and internationally, has put forward various solutions to prevent responsibility gaps in the face of algorithm-induced harm, including the theorization of a direct form of criminal capacity attributed to AI systems. This highlights the urgent need for legislative intervention capable of balancing the competing interests at stake, while ensuring the constitutional principle of criminal law as an exceptional and residual measure (*extrema ratio*).