**VINCENZO MONGILLO**

# Responsabilità da reato degli enti e crimini connessi all'intelligenza artificiale: tecniche giuridiche di intervento e principali ostacoli*

L'utilizzo dell'intelligenza artificiale nelle imprese può assicurare benefici in termini di rafforzamento del sistema di *compliance* penale eventualmente adottato, ma può anche generare pericoli o eventi lesivi penalmente significativi. Questo *Special Report*, presentato in vista del XXI Congresso Internazionale di Diritto Penale dell'AIDP, che si terrà a Parigi dal 26 al 28 giugno 2024 sul tema "Artificial Intelligence and Criminal Justice", si concentra sul secondo profilo d'indagine, dapprima esplorando le possibilità che una persona giuridica possa essere ritenuta responsabile, in base all'attuale quadro normativo italiano e internazionale, per reati collegati all'utilizzo di dispositivi d'intelligenza artificiale, per poi analizzare le tecniche fruibili per sanzionare le aziende in caso di non corretto utilizzo di tali strumenti, nonché le obiezioni e i principali ostacoli da superare.

*Corporate Criminal Liability for AI-Related Crimes: Possible Legal Techniques and Obstacles.*

*The use of Artificial Intelligence in corporations can lead to significant benefits in terms of strengthening the criminal compliance system, but it can also entail significant risks when the actions of those digital tools could cause danger or damage that may result in a criminal offence. This Special Report – presented in view of the AIDP XXI International Congress of Penal Law, which will appear from 26 to 28 June 2024 on the topic 'Artificial Intelligence and Criminal Justice', will focus on the latter research perspective, first exploring the possibility that a legal entity may be held accountable for AI-related crimes by looking at the current regulatory framework, and then investigating possible techniques to make corporations criminally liable in this field, as well as potential counter arguments to such a regulatory objective.*



1. *Introduzione*. Algoritmi ipersofisticati, veicoli senza conducente, robot "umanoidi", sistemi d'arma autonomi e letali, e altro ancora, non sono più solo materiale per sceneggiatori e autori di libri di fantascienza. Siamo, in effetti, entrati nell'era dell'intelligenza artificiale (di seguito anche "IA"), che verosimilmente diventerà la più importante tecnologia del XXI secolo, grazie agli straordinari progressi indotti dall'aumento esponenziale dei dati digitali e

delle capacità computazionali[1].

In ragione di queste epocali conquiste tecnico-scientifiche, attenti osservatori intravedono una quarta rivoluzione industriale[2] e, in prospettiva, «il più grande evento nella storia della nostra civiltà»[3]. Le ricadute potenziali riguarderanno – e in parte già hanno investito – ogni campo della vita sociale, in contesti di pace o di guerra, di lavoro o di svago: medicina, industria, finanza, circolazione stradale, assistenza, riscossione tributaria, intrattenimento, funzionalità delle abitazioni (c.d. domotica), conflitti armati, ecc.[4].

Pertanto, i sistemi giuridici sono chiamati a confrontarsi con le ambivalenti ricadute delle macchine c.d. intelligenti di ultima generazione, che da un lato rappresentano un fattore di miglioramento delle condizioni di vita di intere nazioni e popoli, potendo favorire la crescita umana e sociale, e dall'altro costituiscono un vettore di rischi virtualmente smisurati.

Per queste ragioni, l'IA è divenuto un campo di ricerca assai stimolante anche per i cultori delle discipline penalistiche, come rivela una già considerevole mole di ricerche e di studi. In quest'ambito, si stagliano le indagini sulla responsabilità penale o "da reato" degli enti collettivi, e segnatamente delle società commerciali, tra i maggiori fruitori delle nuove tecnologie. Ciò per un duplice ordine di ragioni, riguardanti tanto il profilo della prevenzione dei reati quanto l'aspetto della repressione.

La prima ragione è che le tecnologie dell'intelligenza artificiale, in combina-

---

* Il testo riproduce, in lingua italiana e con aggiornamenti e integrazioni, lo *Special Report* presentato al Colloquio tenutosi il 15-16 settembre 2022 presso *The Siracusa International Institute for Criminal Justice and Human Rights*, in preparazione del XXI Congresso dell'Associazione Internazionale di Diritto Penale (AIDP), che si terrà a Parigi dal 26 al 28 giugno 2024 sul tema "Artificial Intelligence and Criminal Justice". Una prima versione dell'articolo è stata pubblicata, in inglese, nel volume collettaneo *Traditional Criminal Law Categories and AI: Crisis or Palingenesis*, a cura di Picotti-Panattoni, n. 1/2023 della *Revue Internationale de Droit Pénal*.

[1] Cfr., nella dottrina penalistica, per tutti, BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in *Diritto penale e intelligenza artificiale. "Nuovi scenari"*, a cura di Balbi-De Simone-Esposito-Manacorda, Torino, 2022, 3 s. Semplificando all'estremo, possiamo definire l'IA come qualsiasi macchina in grado di svolgere specifici compiti tipicamente associati all'intelletto umano, mediante algoritmi, programmi informatici e sistemi elettronici.

[2] SCHWAB, *La quarta rivoluzione industriale*, Milano, 2016, *passim*.

[3] Secondo l'opinione del grande fisico, cosmologo e matematico Stephen Hawking, scomparso nel 2018: cfr. *AI will either 'transform or destroy' society, says Prof Stephen Hawking at intelligence centre launch*, in *www.cambridge-news.co.uk*.

[4] Su alcune di queste applicazioni e le ricadute giuridiche, v. i contributi riportati in *Automazione, Diritto e Responsabilità*, a cura di Picotti, Napoli, 2023, spec. 213 ss., 293 ss.

zione con la c.d. *blockchain*, probabilmente cambieranno radicalmente le metodiche e le pratiche della *compliance* penale e la progettazione dei sistemi di controllo interno alle imprese. Nel contesto generale di ciò che può essere designato come il "mondo RegTech"[5], le aziende, utilizzando sofisticati sistemi informatici, possono far emergere *red flags* che altrimenti non sarebbero in grado di identificare attraverso le classiche tecniche di analisi e monitoraggio. Inoltre, possono costruire sistemi decisionali trasparenti e affidabili, in cui è più complicato occultare attività illecite[6].

Ovviamente, quanto precede appartiene al mondo delle aspirazioni; la realtà potrebbe essere meno idilliaca, poiché la nuova frontiera della *compliance* digitale presenta anche forti incognite e rischi da gestire. Tra le questioni che reclamano attenzione, possiamo menzionare, ad esempio: la qualità dei dati alla base dell'elaborazione dei sistemi di intelligenza artificiale[7]; l'impatto sulla protezione dei dati personali e dei diritti dei lavoratori, non senza implicazioni di carattere etico; il ruolo della *cybersecurity* nella salvaguardia dei dati sensibili trattati nei processi di *digital compliance*; le disparità nell'accesso alle risorse digitali; gli effetti macroeconomici sull'occupazione e

[5] Il termine indica la "tecnologia normativa", cioè l'uso della tecnologia digitale per ottimizzare attività di gestione delle prescrizioni normative ed efficientare processi di conformità e controlli operativi.

[6] Su questo tema, anche per quanto riguarda l'impiego della tecnologia non solo come mezzo di prevenzione del crimine ma anche quale strumento per la commissione di reati, si veda PICOTTI, *New Technologies as Tools for and Means Against Crime: Substantial Aspects*, in *Revue Internationale de Droit Pénal*, 2020, 2, 183. Per una recente panoramica e una più ampia rassegna della letteratura in materia, si veda BIRRITTERI, *Corporate Criminal Liability and New Technologies: Digital Compliance Strategies in the Fight against Economic Crimes*, in Aa.Vv., *The Role of Technology in Preventing and Combating Organized Crime, Financial Crimes and Corruption – Book of Proceedings*, OSCE, 2023, 11. Cfr. anche, in generale, SABIA, *Artificial Intelligence and Environmental Criminal Compliance*, in *Revue Internationale de Droit Pénal*, 2020, 1, 179 ss.; SEVERINO, *The Importance of Corporate Compliance in the Digital Era*, in *Revue Internationale de Droit Pénal*, 2021, 2, 435 ss.; MONGILLO, *Presente e futuro della* compliance *penale*, in *www.sistemapenale.it*, 11 gennaio 2022; GULLO, voce *Compliance*, in *Studi in onore di Carlo Enrico Paliero*, a cura di Piergallini-Mannozzi-Sotis-Perini-Scoletta-Consulich, Milano, 2022, 1289 ss.

[7] In un convegno svoltosi il 22 giugno presso l'Università UnitelmaSapienza, il direttore dell'UIF Enzo Serata ha fatto notare alcune criticità, osservando come l'integrale affidamento, nel settore privato, del monitoraggio delle operazioni ad algoritmi intelligenti rischi di provocare un impoverimento della qualità complessiva delle segnalazioni di operazioni sospette (S.O.S.), sicché la valutazione dell'individuo segnalante resta essenziale per assicurare un'informazione di qualità ed evitare di generare un effetto distorsivo, rendendo più complesso lo svolgimento dei compiti assegnati all'UIF. Sulla *compliance* predittiva, anche in tema di antiriciclaggio, cfr. ESPOSITO, *Note sparse sull'intelligenza artificiale*, in *Diritto penale e intelligenza artificiale*, cit., 42-47.

sull'organizzazione del lavoro, che richiedono un'attenta considerazione delle implicazioni sociali e politiche. Ne consegue che, anche sul piano della "conformità" legale, la regolamentazione e la gestione degli strumenti di IA postulano un approccio olistico che tenga conto di aspetti tecnologici, etici, giuridici, di sicurezza e sociali.

La seconda ragione dello spiccato interesse che gli strumenti di IA oggi suscitano nel giurista, incluso il penalista, è ancora più pregnante nell'ottica dell'impatto sociale: il loro utilizzo nel contesto di un'impresa può causare pericoli o danni di varia natura, significativi dal punto di vista della commissione di reati (dolosi o colposi). Si pensi ai settori, già di grande rilievo, delle auto a guida autonoma (*self-driving cars*)[8], della robotica e dei sistemi diagnostici utilizzati in medicina[9], del *trading* finanziario o della gestione logistica mediante algoritmi.

Il problema, come accennato poc'anzi, è amplificato dal seguente dato fattuale: gran parte dei sistemi di IA sono prodotti o utilizzati da persone giuridiche e in particolare da imprese societarie. A questo riguardo, la questione giuridica chiave è se gli enti collettivi coinvolti possano *de iure condito* o debbano *de iure condendo* essere chiamati a rispondere in sede penale o para-penale (nell'ordinamento italiano ai sensi del d.lgs. 8 giugno 2001, n. 231) per reati collegati all'utilizzo di IA.

Nel presente lavoro intendiamo concentrarci su quest'ultima prospettiva, esplorando, innanzitutto, la possibilità che un'organizzazione possa essere ritenuta responsabile per la commissione di reati in vario modo vincolati all'impiego di apparecchi o algoritmi di IA. Esamineremo, quindi, l'attuale quadro normativo[10], scandaglieremo le tecniche che paiono in astratto sperimentabili per responsabilizzare gli enti in queste sfere peculiari di rischio, e metteremo a fuoco i possibili ostacoli e le argomentazioni che si oppongono a

[8] A livello monografico, esaustivamente, LANZI, Self-driving cars *e responsabilità penale. La gestione del "rischio stradale" nell'era dell'intelligenza artificiale*, Torino, 2023.

[9] Si pensi ai robot già da tempo impiegati nelle operazioni chirurgiche, agli algoritmi diagnostico-terapeutici, ecc. Su quest'inedita dimensione di rischio, cfr., da ultimo, l'indagine di AMORE-ROSSERO, *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità. Una ricerca sociologica e giuridico penale*, Bologna, 2023, e in particolare i capitoli di taglio penalistico redatti da AMORE, ivi, 101 ss., 145 ss. e 177 ss.

[10] V. par. 2.

tale approdo giuridico[11]. Infine, tireremo le fila delle considerazioni svolte[12].

2. *I regimi di responsabilità penale (o para-penale) dell'impresa a livello internazionale e i crimini legati all'IA: lo scenario attuale*. Quando si affronta il dilemma se una persona giuridica possa essere dichiarata responsabile per un reato caratterizzato, nella fase della preparazione o in quella della realizzazione, dall'impiego di strumenti di IA, occorre distinguere tra sistemi informatici non autonomi e sistemi che possono prendere decisioni interamente autonome, che gli stessi programmatori o utenti finali non sono in grado, in tutto o in parte, di prevedere[13].

2.1. *Sistemi di IA non autonomi e sistemi autonomi deliberatamente utilizzati per commettere reati*. Partendo dalla prima costellazione casistica, i sistemi *non autonomi* non sembrano sollevare problemi aggiuntivi rispetto a quelli tradizionalmente affrontati nel diritto penale delle persone fisiche o nel diritto punitivo degli enti pluripersonali.

Tali dispositivi, infatti, si limitano a eseguire le istruzioni ricevute dal programmatore umano o, comunque, operano sotto il diretto controllo di uno o più soggetti individuali. Di conseguenza, possono essere attivati gli "ordinari" canali di *enforcement* e di imputazione delle responsabilità penali all'individuo ed eventualmente anche alla persona giuridica[14].

Infatti, sia nei casi in cui lo strumento tecnologico venga utilizzato al precipuo scopo di commettere un reato contro la vita, l'incolumità, la riservatezza personale, il patrimonio altrui, ecc., sia nelle situazioni in cui l'uso del sistema provochi un danno non voluto (ossia ascrivibile a colpa) e comunque in grado di integrare un fatto penalmente illecito[15], la responsabilità potrà essere invo-

[11] V. par. 3.
[12] V. par. 4.
[13] Per una panoramica generale v. S. BECK, *The Problem of Ascribing Legal Responsibility in the Case of Robotics*, in 31 *AI & Soc.*, 2016, 473 ss.
[14] Su questo tema si veda anche PANATTONI, *AI and Criminal Law: the Myth of 'Control' in a Data-Driven Society*, in *Revue Internationale de Droit Pénal*, 2021, 1, 125 ss.
[15] CALDWELL *et al.*, *AI-enabled future crime*, in 9 *Crime Science*, 2020, 14, ha fatto una rassegna dei possibili usi dei sistemi di IA nella perpetrazione di reati. Nella dottrina italiana, v., tra gli altri, MAGRO, *A.I.: la responsabilità penale per la progettazione, la costruzione e l'uso dei robot*, in *il Quot. giur.*, 12 giugno 2018.

cata in base ai principi e alle regole consueti nei diversi ordinamenti. Dell'illecito penale potranno essere chiamati a rispondere – a seconda dei casi – il produttore, il distributore o l'utente finale del sistema[16], nonché la persona giuridica o le persone giuridiche volta per volta coinvolte. Rispetto all'ente collettivo, i modelli di imputazione spaziano dalle forme basate sulla sulla *vicarious liability* o sulla *strict liability* (responsabilità oggettiva), a quelle centrate sulla colpa di organizzazione (*organisational fault*) o sulla mancata prevenzione di specifici reati (*failure to prevent model*)[17].

Ovviamente, anche tra questi casi più "semplici" possono sorgere, in concreto, delicate questioni di attribuzione della responsabilità, dal punto di vista sia dell'autore o concorrente individuale sia dell'organizzazione di appartenenza. Tendenzialmente, come si è accennato, queste difficoltà applicative non si discostano da quelle normalmente affrontate in materia di responsabilità da prodotto difettoso[18]. Tuttavia, i tradizionali problemi di imputazione sono acuiti, nel contesto che ci riguarda, dall'incidenza sulla dinamica fattuale di tecnologie complesse come quelle dell'IA, avuto riguardo, soprattutto, ai metodi utilizzati nella fabbricazione e alla marcata complessità tecnica dell'*output* produttivo.

In particolare, dall'angolazione della responsabilità individuale, risultano amplificate – per menzionare solo quelle più vistose – le questioni dommatiche concernenti l'identificazione dei responsabili e delle cause penalmente rilevanti dell'evento, la distinzione tra azione e omissione, le fonti legali e la delimitazione delle posizioni di garanzia dei vari attori del processo di produzione, programmazione e commercializzazione della macchina, la prevedibilità *ex ante* dell'eventuale difettosità del prodotto, i parametri della cooperazione colposa e del concorso di cause colpose indipendenti, il ruolo del principio di affidamento.

Pesa anche la tendenziale distribuzione del processo produttivo in *supply*

[16] Nella letteratura italiana, v. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 2020, 1753.

[17] Sul punto v. anche FE. MAZZACUVA, *The Impact of AI on Corporate Criminal Liability: Algorithmic Misconduct in the Prism of Derivative and Holistic Theories*, in *Revue Internationale de Droit Pénal*, 2021, 1, 143 ss.

[18] PIERGALLINI, *Intelligenza artificiale*, cit., 1753.

*chain* estremamente frammentate: un aspetto che non contraddistingue solo l'industria dell'IA, ma che qui assume tratti esasperati. La figura del "produttore" si frantuma in una miriade di operatori economici, distribuiti a livello nazionale e globale, per ciascuna componente del bene finale. Ciò complica la ricostruzione delle posizioni di garanzia e degli anelli causali, e amplifica le difficoltà di previsione degli effetti dell'insieme da parte dei singoli partecipanti alla catena eziologica. Ma anche nel caso in cui il processo produttivo sia concentrato in un'unica organizzazione, l'immagine monolitica del "fabbricante" o del "progettista" corrisponde a un utilizzo sincopato del linguaggio, posto che nei processi di costruzione delle macchine intelligenti intervengono una moltitudine di tecnici che forniscono il loro apporto al progetto unitario.

Possono profilarsi, inoltre, problemi di giurisdizione, dovuti alla scissione planetaria dei diversi segmenti produttivi[19]. Sotto quest'ultimo profilo, anche il rapporto tra venditore e consumatore può risultare profondamente divaricato sul piano geografico. Basti considerare che chiunque oggi può acquistare *on line* strumenti di IA – ad esempio droni – da venditori residenti in Paesi terzi.

Circa, specificamente, i profili di possibile imputazione della responsabilità a un ente collettivo, vanno tenute in conto anche le peculiarità dei diversi regimi normativi. Si consideri, per esempio, il caso in cui la messa su strada di un'auto a guida autonoma sotto la supervisione di un conducente umano porti alla realizzazione di un omicidio colposo e, tuttavia, le norme applicabili in materia di responsabilità dell'impresa non includano tale illecito tra quelli in grado di far scattare la responsabilità della *societas* (c.d. reati-presupposto). Evidentemente, questo problema applicativo può sorgere solo nei sistemi nazionali basati su un "catalogo chiuso" (*numerus clausus*) di reati-matrice della responsabilità della *corporation*[20].

---

[19] In merito all'impatto della globalizzazione economica sull'applicazione della legge penale nello spazio, sia permesso il rinvio a MONGILLO, Forced labour *e sfruttamento lavorativo nella catena di fornitura delle imprese: strategie globali di prevenzione e repressione*, in *Riv. trim. dir. pen. econ.*, 2019, 3-4, 630 ss., spec. par. 6; ID., *Criminalità di impresa transfrontaliera e giurisdizione penale "sconfinata": il difficile equilibrio tra efficienza e garanzie*, in *Riv. it. dir. proc. pen.*, 2023, 1, 111 ss.

[20] Come accade con il nostro d.lgs. 8 giugno 2001, n. 231: per una panoramica (in lingua inglese) si veda DE MAGLIE, Societas Delinquere Potest? *The Italian Solution*, in *Corporate Criminal Liability. Emergence, Convergence and Risk,* a cura di Pieth-Ivory, Dordrecht, 2011, 255 ss.

Gli esiti di questo primo snodo della nostra analisi possono essere replicati senza soverchie difficoltà anche rispetto ai *sistemi di IA completamente autonomi* ma sin dall'origine congegnati con l'*obiettivo di commettere reati* e causare intenzionalmente danni a terzi, ad esempio per scopi terroristici o di destabilizzazione di governi legittimi.
In questi casi, gli applicativi di IA si pongono come una sorta di *longa manus* di chi intende perpetrare fatti di reato[21]. Ne consegue che la responsabilità penale potrà essere attribuita – come sottolineato in precedenza – alle persone fisiche che utilizzano lo strumento con propositi criminali, nonché alla persona giuridica di appartenenza, quando le regole vigenti lo consentano.
Inoltre, in caso di condanna dell'autore individuale e/o dell'ente, il congegno dannoso potrà essere, di norma, confiscato come *instrumentum sceleris*, vale a dire come strumento adoperato per commettere il reato, nel nostro ordinamento ai sensi dell'art. 240 c.p.

2.2. *Sistemi di IA completamente autonomi e non programmati/utilizzati per commettere reati*. Come anticipato, quelli appena passati in rassegna sono i casi più semplici, gli *easy cases* direbbe Hart.
La questione diventa molto più nebulosa e complessa in relazione a sistemi di IA *completamente autonomi* e *non programmati* o impiegati *per realizzare reati*. Questo sono i veri *hard cases*, quelli che impegneranno maggiormente i tribunali non appena gli scenari che oggi paventiamo cominceranno a divenire una possibilità più realistica.
Ci riferiamo agli strumenti digitali strutturati *by design* per apprendere automaticamente e agire "*alone*", quindi in grado di percepire l'ambiente circostante, interagire con esso, analizzare i dati, fare previsioni, prendere decisioni e innescare modificazioni della realtà esterna in maniera del tutto indipendente sia dal produttore del sistema sia dall'utente[22].
Nella letteratura scientifica si parla di *machine learning*, la cui ultima evolu-

[21] Su questo tema, cfr. anche SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 2021, 1, 83 ss.
[22] Sul punto, dalla prospettiva specifica delle armi autonome e delle loro implicazioni in materia di diritto penale, v. CROOTOF, *War Torts: Accountability for Autonomous Weapons*, in 164 *University of Pennsylvania Law Review*, 2016, 1347 ss.

zione sono le sofisticate tecniche di *deep learning*[23]. La macchina, in questo caso, è un sistema aperto[24] che apprende continuamente e automaticamente, con conseguente modificazione delle connessioni tra i neuroni artificiali, sicché anche sotto questo profilo l'ambizione è imitare, fin dove possibile, la "plasticità" del cervello umano e l'incessante mutamento delle reti neurali che lo compongono[25].

Dalla prospettiva penalistica, il problema più delicato generato da questi sviluppi della tecnica è proprio l'impossibilità di pronosticare, almeno per intero, il funzionamento futuro del sistema, *ergo* tutte le decisioni che, nelle infinite situazioni della vita reale, l'apparecchio di IA potrà assumere a prescindere dall'istruzione o dall'autorizzazione di una guida umana.

Se le macchine intelligenti fossero interamente automatizzate, questo problema non si porrebbe, dato che per definizione si può automatizzare tutto ciò che è prevedibile. Al contrario, i dispositivi di ultima generazione capaci di autoapprendimento assumono decisioni attraverso il contatto con l'ambiente esterno e i dati immagazzinati nel *cloud*, «un potente *hub* computazionale in grado di conservare, elaborare ed erogare enormi masse di dati[26], da cui attingere continuamente per gli aggiornamenti, gli *upgrading*»[27]. La conseguenza è che né il progettista, né il programmatore, né l'utente finale possono conoscere esattamente e in anticipo il *pattern* comportamentale che la macchina sceglierà, volta per volta, nell'interpretare le infinite situazioni della vita reale[28].

Per certi aspetti, una dose di imprevedibilità è finanche preordinata, giacché l'obiettivo del produttore di questi dispositivi così evoluti non è istruire e regolare anticipatamente qualsiasi scelta, ma far sì che la tecnologia "cogitante" lavori e assuma le sue decisioni in un modo che si confida sarà il più efficace

[23] Cfr. SEJNOWSKI, *The Deep Learning Revolution*, Cambridge, 2018.

[24] V. anche MAGRO, Robot*,* cyborg *e intelligenze artificiali*, in *Cybercrime*, in *Omnia. Trattati giuridici*, diretto da Cadoppi-Canestrari-Manna-Papa, Torino, 2019, 1191.

[25] Le reti neurali artificiali sono «composte da elementi fra loro collegati, che lavorano in sincrono sul modello dei neuroni biologici e delle loro sinapsi»: così, BODEI, *Dominio e sottomissione. Schiavi, animali, macchine e Intelligenza Artificiale*, Bologna, 2019, 317.

[26] Si parla, al riguardo, di *big data*.

[27] BODEI, *Dominio e sottomissione*, cit., 315.

[28] Sulla questione dell'imprevedibilità tecnologica delle "macchine intelligenti", cfr., ad es., CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, in *Diritto penale e intelligenza artificiale*, cit., 23-25.

possibile. Questa è anche l'unica declinazione ipotizzabile del principio di affidamento nel rapporto uomo-macchina intelligente. Ma il principale fattore di novità, rispetto alle classiche declinazioni del principio di affidamento nella teoria della colpa penale, è che nella specie non si tratta di fare assegnamento su una persona umana che si abbia ragione di ritenere sufficientemente esperta, addestrata e avveduta per assolvere determinati compiti, ma su una *learning machine*, ipoteticamente disposta a un continuo percorso esperienziale di apprendimento e affinamento delle sue "abilità". Qui, come detto, giace il *punctum dolens* penalistico delle tecnologie in discorso.

È evidente allora come, attraverso quest'inedita dimensione empirica, si passi dal *pericolo prevedibile* e in tutto o in parte quantificabile, che rappresenta il dominio della prevenzione, al *rischio ignoto*, che si colloca, elettivamente, nel campo della c.d. precauzione. Difatti, l'idea di *prevenzione* è contrassegnata dal sapere scientificamente corroborato. Il *principio di precauzione*, invece, è legato agli ambiti di rischio segnati da una consistente incertezza scientifica, sicché riguarda ciò che si sospetta possa accadere ma non si sa se – e in che termini – accadrà.

In una società che tende a rifiutare l'idea del "fortuito", è facile pronosticare una vibrante domanda di giustizia delle vittime a fronte di qualsiasi fallimento dell'algoritmo auspicabilmente "intelligente". Tuttavia, secondo la prevalente opinione dottrinale, non può ammettersi il ricorso al principio di precauzione quale fonte giuridica integrativa del dovere di diligenza, per supportare imputazioni personali a titolo di colpa penale. Infatti, a rigore, questo criterio d'imputazione soggettiva «richiede, anzitutto, la violazione di regole cautelari a fondamento nomologico, orientate alla prevenzione di eventi prevedibili (e non soltanto ipotizzabili) *ex ante*: sulla base, quindi, del patrimonio cognitivo disponibile (per l'agente modello di riferimento o almeno per l'agente concreto per avventura dotato di conoscenze superiori) al momento della condotta, la cui antidoverosità non può essere recuperata in maniera retroattiva»[29].

La *prospettiva precauzionale* si risolve, così, in un metodo di buona gestione amministrativa del rischio ipotetico, che può suggerire, in casi limite, anche il

[29] CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, Roma, 2012, 47.

ricorso a divieti assoluti di svolgere attività o utilizzare tecnologie di cui si teme la capacità di generare pericoli non ancora avvalorati scientificamente, ma che appaiono prospetticamente insostenibili per gravità e diffusività.

Fissati i paletti concettuali della nostra riflessione, non si può escludere che le decisioni autonomamente assunte da un sistema di IA possano corrispondere, perlomeno da un punto di vista materiale-oggettivo, a una condotta penalmente tipica. Vengono in mente, ad esempio, abusi di mercato realizzati mediante un *software* di IA capace di gestire in autonomia operazioni borsistiche o transazioni finanziarie[30], oppure danni all'integrità fisica – omicidio o lesioni in ipotesi colposi – causati da robot/sistemi di IA parimenti dotati di autonomia operativa[31].

A volte la sperimentazione sul campo di una certa macchina (ad es. un dispositivo diagnostico o utilizzato in interventi chirurgici) oppure i *feedback* dei clienti potrebbero consentire al produttore e/o all'utilizzatore di acquisire la consapevolezza che una determinata tipologia di IA devierà, in una certa percentuale di casi (anche non esattamente quantificabile), dalle funzioni programmate, innescando processi eziologici potenzialmente dannosi e che non si sia in grado di neutralizzare, con apposite misure, allo stato delle cognizioni scientifiche. In queste evenienze, possiamo parlare di un rischio minimo *noto*, che potrà spingere l'autorità statale alla decisione giuspolitica di tollerarlo, quando l'effetto collaterale avverso appaia di scarso rilievo, oppure di vietarlo quale rischio giuridicamente non consentito.

Tuttavia, nello scenario caratteristico del *machine learning* e, viepiù, del *deep learning*, il sapere scientifico-esperienziale, di norma, non consente né di affermare che l'utilizzo di un certo sistema di IA potrebbe causare specifici pericoli o danni non controllabili, né di escluderli categoricamente.

---

[30] Si veda CONSULICH, *Il nastro di Mobius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca, borsa e titoli di credito*, 2018, 2, 195 ss.; PALMISANO, *L'abuso di mercato nell'era delle nuove tecnologie.* Trading *algoritmico e principio di personalità dell'illecito penale*, in *Dir. pen. cont.*, 2019, 2, 129 ss.; TRIPODI, *Abusi di mercato e* trading *algoritmico*, in *Il diritto nell'era digitale. Persona, Mercato, Amministrazione*, a cura di Giordano-Panzarola-Police-Preziosi-Proto, Milano, 2022, 745 ss.

[31] V., ad es., CROOTOF, *War Torts*, cit., 1347 ss.; nonchè, KING-AGGARWAL-TADDEO-FLORIDI, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in 26 *Science and Engineering Ethic*s, 2020, 89 ss.

La responsabilità penale individuale, in simili scenari, è esposta a grandi incertezze e a dubbi tendenzialmente irresolubili. Ma anche la possibilità o meno di responsabilizzare in sede penale le organizzazioni societarie eventualmente implicate espone il giurista alla tenaglia di un dilemma del quale è difficile venire a capo con soluzioni univoche.

Prima di rispondere a quest'ultimo interrogativo, ci sembra necessario sgombrare il campo da un'ipotesi fantasiosa, che pure aleggia – e di tanto in tanto si affaccia – nel dibattito scientifico: l'idea cioè di attribuire la *personalità giuridica* ai sistemi di intelligenza artificiale in quanto tali e *punire direttamente la macchina*[32], in presenza di un cattivo funzionamento tale da generare pregiudizi a interessi giuridicamente protetti, significativi dal punto di vista penale.

Tale suggestione evoca i processi medievali contro gli animali o finanche l'animismo dei popoli primitivi[33], vale a dire il fenomeno consistente nell'attribuire proprietà spirituali a realtà fisico-materiali.

È evidente, però, come la soluzione ai problemi giuridici qui affrontati non possa venire, nella modernità iper-tecnologica che ci circonda, dalla riesumazione del pensiero arcaico più irrazionale e quindi dal regresso – a dire il vero

[32] V., per tutti, il pensiero del penalista israeliano HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Dordrecth, 2015, che sostiene la tesi della responsabilità penale diretta dell'IA; ID., *The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 2010, vol. IV, 171 ss. *Contra*, ad es., 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*, *www.ejc.reeps.com*, 2017, 9, il quale – in senso contrario al punto di vista espresso da SÁNCHEZ DEL CAMPO REDONET, *Cuestiones jurídicas que plantean los robots*, in *Revista de privacidad y derecho digital*, 2016, 2 – osserva che è impossibile considerare i robot capaci di commettere reati e la distruzione fisica della macchina come l'equivalente della pena di morte, dal momento che condizione ontologica del diritto penale è che le sue possibili reazioni siano conosciute o conoscibili *ex ante* dagli ipotetici futuri trasgressori, al pari dei comandi e dei divieti penali (c.d. conoscibilità del diritto). L'A. conclude notando come rifiutare "la responsabilità penale della macchina" non equivalga a irrilevanza di quello che "fa" la macchina (p. 10). Sulla questione, v. anche BASILE, *La responsabilità penale dei sistemi di intelligenza artificiale: scienza o fantascienza?*, in *Automazione*, cit., 103 ss.; PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, in *Il diritto nell'era digitale*, cit., 722 ss.

[33] Cfr. KELSEN, *Dottrina pura del diritto*[3] (1960), Torino, trad. it., 1966, 42 s., 102 s.; ID., *Teoria generale del diritto e dello Stato* (1945), trad. it., Milano, 1994, 94, 195 ss., sull'interpretazione animistica della natura dei popoli primitivi (secondo tale interpretazione «si crede che ogni oggetto del mondo percettivo sia la dimora di uno spirito invisibile che è il signore dell'oggetto, che «ha» l'oggetto, nella stessa guisa in cui la sostanza ha le sue qualità, o il soggetto grammaticale i suoi predicati»); il fenomeno antropologico animico è stato accuratamente indagato, in chiave psicoanalitica, anche da FREUD, *Totem e tabù*, trad. it., Milano, 1989, 89 ss.

grottesco – alle fasi primitive dello sviluppo sociale; e in effetti, ad oggi, nessuna legislazione (almeno a noi nota) prevede misure o sanzioni penali direttamente applicabili a strumenti informatici o congegni di IA[34].

Allo stato delle conoscenze tecnico-scientifiche, v'è unanime consenso circa il fatto che questi strumenti asseritamente intelligenti non abbiano alcuna capacità di autodeterminazione, e quindi una vera e propria *self-consciousness* o identità personale, intesa – quantomeno a partire dal padre dell'empirismo moderno John Locke – come la coscienza che una persona abbia del suo permanere attraverso il tempo e le fratture dell'esperienza[35]. L'autocoscienza è indispensabile per fondare una responsabilità penale *stricto sensu*, e da essa dipende anche la possibilità che una qualunque sanzione punitiva (a prescindere dalla sua qualificazione) possa motivare psicologicamente il destinatario al rispetto della norma. In breve, un dispositivo di IA, «per quanto "intelligente", rimane pur sempre una macchina»[36]. Lo stesso utilizzo del termine "intelligenza" riflette un linguaggio metaforico con cui attribuiamo all'apparecchio qualità di cui è, nella realtà, privo[37].

---

[34] Si veda il *General Report* di PICOTTI, in *Traditional Criminal Law Categories and AI*, cit., 11 ss., basato anche sui "rapporti speciali" e sull'analisi delle risposte dei gruppi nazionali al questionario elaborato dall'AIDP, in cui l'A. sottolinea che «In all the countries of the national reports collected, AI systems do not have legal personhood or legal capacity. Therefore, they cannot be considered as a legal entity»; pertanto, «they cannot be considered as subject of criminal law either». Questo problema è stato affrontato nella letteratura scientifica in materia già da diversi anni, a partire dagli studi pionieristici di SOLUM, *Legal Personhood for Artificial Intelligences*, in 70 *North Carolina Law Review*, 1992, 1231 ss.

[35] In effetti, tale approccio alla "personalità", nel pensiero filosofico, risale almeno a LOCKE, *An Essay Concerning Human Understanding,* London, 1690, trad. it. della 4ª ed., 1699, *Saggio sull'intelletto umano*, Torino, 1971, lib. II, cap. XXVII, spec. n. 11, 394 s. («poiché la coscienza accompagna sempre il pensare ed è ciò che fa sì che ognuno sia quello che egli chiama io, distinguendo con ciò se stesso da tutti gli altri esseri pensanti, in questo solo consiste l'identità personale, cioè nel fatto che un essere razionale è sempre lo stesso. E fin dove questa coscienza può essere estesa indietro ad una qualsiasi azione o pensiero del passato, fin lì giunge l'identità di quella persona»); n. 20, p. 402 («Su questa identità personale è fondato tutto il diritto e la giustizia della ricompensa e del castigo»); n. 28, p. 407 («Solo mediante la coscienza la personalità [...] riconosce come sue e imputa a se stessa azioni passate sulla stessa base e per la stessa ragione per cui lo fa per azioni presenti»).

[36] Messaggio del Santo Padre Francesco per la 57ma Giornata Mondiale della Pace (1° gennaio 2024), 14.12.2023, in cui si osserva che i sistemi digitali c.d. intelligenti «sono, in ultima analisi, "frammentari", nel senso che possono solo imitare o riprodurre alcune funzioni dell'intelligenza umana» e, con specifico riferimento ai sistemi d'arma autonomi, si nota come essi «non potranno mai essere soggetti moralmente responsabili: l'esclusiva capacità umana di giudizio morale e di decisione etica è più di un complesso insieme di algoritmi, e tale capacità non può essere ridotta alla programmazione di una macchina [...]».

[37] BODEI, *Dominio e sottomissione*, cit., 300.

Si potrebbe provare ad azzardare la seguente replica: moltissimi ordinamenti nazionali già ammettono una responsabilità penale a carico delle persone giuridiche, di per sé prive di consistenza psico-fisica[38]. Il paragone, ovviamente, non regge: un algoritmo non è solo privo di *self-consciousness*, ma pure del sostrato proprio di una comunità umana. Di contro, una *legal person* non è solo astrazione giuridica, ma organizzazione di persone e di mezzi e dunque anche collettività di individui in carne ed ossa (come gli amministratori, i dipendenti e i collaboratori di una società commerciale), questi sì motivabili da un precetto normativo e dalla minaccia sanzionatoria per l'inosservanza.

Resta quindi da affrontare la questione se, a fronte di reati la cui dinamica causale sia stata influenzata dall'uso di strumenti di IA completamente autonomi, possa essere punita anche la persona giuridica che – attraverso suoi membri – a quella tecnologia abbia fatto ricorso, oppure che abbia progettato, prodotto, distribuito o venduto il dispositivo.

A tal fine sono necessari alcuni chiarimenti preliminari.

In primo luogo, vale la pena di considerare che, a livello internazionale, la responsabilità penale delle società è, di norma, indissolubilmente legata alla commissione di un reato da parte di un individuo (il cosiddetto "fatto di connessione"; in tedesco *Anknüpfungstat*)[39]. Tuttavia, nei casi sopra menzionati, è arduo – e il più delle volte impossibile – affermare la responsabilità penale di una persona fisica.

Può risultare proibitiva già la dimostrazione giudiziale dell'esistenza della fattispecie oggettiva del reato (in inglese, *actus reus*), ad esempio per ciò che attiene al nesso causale tra l'evento terminale offensivo e un errore di progetta-

[38] Così, in effetti, GLESS-SILVERMAN-WEIGEND, *If Robots Cause Harm, Who Is To Blame? Self-Driving Cars and Criminal Liability*, in 19 *New Criminal Law Review*, 2016, 3, 412 ss.

[39] Si vedano, tra gli altri, *European Developments in Corporate Criminal Liability*, a cura di Gobert-Pascal, Oxon-New York, 2011 e DE MAGLIE, *Models of Corporate Criminal Liability in Comparative Law*, in 4 *Wash. U. Global Stud. L. Rev.*, 2005, 547. Per una panoramica dei principali modelli, anche al di fuori dell'Europa, v. PIETH-IVORY, *Emergence and Convergence: Corporate Criminal Liability Principles in Overview*, in *Corporate Criminal Liability*, cit., 3 ss. Nella letteratura italiana, cfr. MONGILLO, *La responsabilità penale tra individuo ed ente collettivo*, Torino, 2018, 175-262; ID., *The Allocation of Responsibility for Criminal Offences between Individuals and Legal Entities in Europe*, in *Corporate Criminal Liability and Compliance Programs*, a cura di Fiorella, Napoli, 2012, vol. II, *Towards a Common Model in the European Union*, 121 ss.; DE SIMONE, *Profili di diritto comparato*, in *Responsabilità da reato degli enti*, a cura di Lattanzi-Severino, Torino, 2020, vol. I, *Diritto sostanziale*, 3 ss.

zione di un singolo individuo, considerato altresì che moltissime persone – e a volte variegate galassie aziendali e multi-aziendali – cooperano per creare e lanciare determinati prodotti sul mercato.

Ma a parte ciò, il vero e insuperabile "ingombro" tende a divenire la prova della fattispecie soggettiva (*mens rea*), posto che, nei casi in questione, non è possibile incolpare un responsabile individuale che certamente non aveva intenzione di commettere un reato e non poteva nemmeno prevedere il comportamento del sistema informatico, salva la semplice e astratta possibilità che qualcosa di negativo potesse andare storto: ma in questo caso si tornerebbe non già al campo della prevenzione, ma a quello della precauzione[40] o, al più, di una generica e inconsistente *prevedibilità della imprevedibilità*. Come abbiamo già avuto modo di rimarcare, infatti, questi dispositivi intelligenti prendono decisioni autonome e, di fatto, non pronosticabili, mentre pure i meccanismi con cui apprendono e compiono determinate scelte risultano sovente sconosciuti (il nodo della c.d. "*black-box*")[41].

In breve, l'impossibilità – vuoi teorica vuoi pratica – di rendere un essere umano penalmente responsabile (a titolo di dolo, colpa o ipotesi intermedie come la *recklessness* inglese), significa che anche le *corporation* non potranno essere responsabilizzate e sanzionate.

Ciò è indubitabilmente vero per i modelli "derivativi" di responsabilità, vale a dire quelli in cui l'ente risponde indirettamente del reato commesso da un *agent*, nell'ambito del rapporto che lo lega all'organizzazione e nell'interesse di questa. Il tradizionale meccanismo imputativo di matrice statunitense, noto

---

[40] Sull'impossibilità di applicare il principio di precauzione per integrare normativamente il reato colposo, con riferimento alla tematica in oggetto, v., nella dottrina spagnola, QUINTERO OLIVARES, *La robótica ante el derecho* penal, cit., 1 ss.

[41] Su questi aspetti, v. GRANDI, *Positive obligations (*Garantestellung*) grounding the criminal responsibility for not having avoided an illegal result connected to the AI functioning*, in *Traditional Criminal Law Categories and AI*, cit., 67 ss. V. anche MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, in *Revue Internationale de Droit Pénal*, 2021, 1, 109 ss.; PAGALLO, *From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability*, in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, IJCAI, 2017; nella letteratura italiana, MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen.*, 2020, 1 ss.; FRAGASSO, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, in *Dir. pen. cont.*, 2023, 1, 31 s. Per una panoramica generale sui legami tra IA e diritto, si vedano i diversi contributi pubblicati in *Regulating Artificial Intelligence*, a cura di Wischmeyer-Rademacher, Cham, 2020.

come *vicarious liability*, implica, infatti, che le *corporation* possano essere ritenute colpevoli e punite solo quando sia possibile identificare una persona fisica che abbia tenuto una condotta in grado di soddisfare tutti i requisiti oggettivi e soggettivi del reato rilevante[42].

Alcuni studiosi[43] hanno riflettuto sull'eventuale estensione della *corporate mind* – su cui fondare un addebito di responsabilità – anche alle disfunzioni algoritmiche co-causatrici di danni a terzi, argomentando la possibilità di immaginare illeciti "commessi" da sistemi digitali e di considerarli illeciti aziendali (*corporate wrongs*), analogamente a quelli perpetrati da dipendenti.

Si tratta, all'evidenza, di costruzioni artificiose, perché raffrontano dati empirici oggettivamente incomparabili. Partendo da questa incongrua assimilazione, un simile approccio pretenderebbe di considerare l'algoritmo come un *agent* dell'organizzazione e, soprattutto, di individuare una colpevolezza/*mens rea* in capo ad essi per poi imputarla alla persona giuridica. L'impostazione è impraticabile, proprio per la già riscontrata impossibilità di cogliere un elemento di *Gewissen* (coscienza morale), o più semplicemente di coscienza di sé, nella macchina.

La risposta al quesito resta negativa anche ipotizzando il ricorso a un diverso modello di responsabilità, basato sulla teoria della identificazione, pure di derivazione anglosassone (*identification doctrine*), sebbene non priva di ascendenze nella teoria pubblicistica dell'organo di matrice continentale[44]. Essa comporterebbe l'ulteriore problema di come concepire l'algoritmo in quanto tale come «directing mind and will of the company»[45].

---

[42] Cfr. NANDA, *Corporate Criminal Liability in the United States: Is a New Approach Warranted?*, in *Corporate Criminal*, cit., 63 ss.; WELLS, *Corporations and Criminal Responsibility*², New York, 2018, *passim*. Per una disamina delle pratiche di *enforcement* negli Stati Uniti in relazione a questo modello di responsabilità penale della *corporation*, v., in una letteratura sterminata, ARLEN, *Corporate Criminal Enforcement in the United States: Using Negotiated Settlements to Turn Corporate Criminals into Corporate Cops*, in 17.12 *NYU School of Law Public Law Research Paper*, 2017, 1 ss.; GARRETT, *Too Big to Jail. How Prosecutors Compromise with Corporations*, Belknap, 2014.

[43] Il riferimento è, in particolare, al lavoro di DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, in 98 *North Carolina Law Review*, 2020, 893. Su questo tema si vedano anche l'analisi e le soluzioni proposte da ABBOT-SARCH, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, in 53 *UC Davis Law Review*, 2019, 323.

[44] MONGILLO, *La responsabilità penale tra individuo ed ente collettivo*, cit., spec. 123 ss.

[45] Si vedano: WELLS, *Corporate Criminal Liability in England and Wales: Past, Present and Future*, in *Corporate Criminal Liability*, cit., 91 ss.; GOBERT, *Corporate Criminality: Four Models of Fault*, in 14

Infine, gli esiti non mutano neppure spostando la nostra lente sugli ordinamenti in cui la responsabilità dell'entità collettiva è fondata su un requisito di *colpa di organizzazione* o, similmente, costruito in termini di omessa prevenzione del reato *(failure to prevent)*. L'idea alla base di questi sistemi è che le organizzazioni possano essere ritenute responsabili per non aver messo in atto *compliance programs* o procedure adeguate a prevenire il verificarsi di specifici reati.

Non s'ignora come, a certe condizioni, molti di questi regimi di responsabilità corporativa ammettano la possibilità di dichiarare la *societas* "autonomamente" responsabile, a prescindere cioè dall'identificazione materiale dell'autore del reato (c.d. "*anonymous guilt*"). Nel paradigma "autonomista" possiamo, senz'altro, collocare l'art. 8 del d.lgs. n. 231/2001 italiano[46]. Tuttavia, pure questi modelli di responsabilità restano ancorati all'esigenza di appurare la commissione di un reato in tutti i suoi elementi essenziali, materiali e soggettivi, o quanto meno i coefficienti oggettivi della condotta umana e dell'evento, al più consentendo all'autorità giudiziaria di prescindere dall'identificazione dello specifico autore individuale[47].

3. *Le prospettive future: responsabilizzare gli enti per i reati collegati all'IA?* È il momento di riflettere, concisamente, sulle tecniche legislative che potrebbero essere adottate, nel futuro dominabile dal nostro sguardo, al fine di fondare la responsabilità di una società per reati determinati dall'utilizzo di sistemi di IA, e in particolare di quelli completamente autonomi. Occorre comprendere, altresì, se le diverse strategie politico-criminali siano o meno appropriate ed eque.

La nostra prima tesi è che non si debba affrontare questo problema da una

---

*Legal Stud.*, 1994, 393.

[46] Sui problemi applicativi che solleva questa disposizione, per un'approfondita trattazione, sia consentito il rinvio a MONGILLO, *La responsabilità penale tra individuo ed ente collettivo*, cit., 310.

[47] Sulle norme che regolano la responsabilità da reato degli enti v., per quanto riguarda il quadro normativo italiano e per una panoramica generale, DE MAGLIE, *Societas Delinquere Potest?*, cit., 255. Con riferimento al modello britannico del *failure to prevent*, cfr., tra gli altri, WELLS, *Corporate Responsibility and Compliance Programs in the United Kingdom*, in *Preventing Corporate Corruption. The Anti-Bribery Compliance Model*, a cura di Manacorda-Centonze-Forti, Cham, 2014; CAMPBELL, *Corporate Liability and the Criminalization of Failure*, in 12 *Law and Financial Markets Review*, 2018, 2, 57 ss.; SULLIVAN, *The Bribery Act 2010: An Overview*, in 2 *Criminal Law Review*, 2011, 87 ss.

prospettiva unilaterale e monistica. Soprattutto in relazione alle nuove tecnologie dell'IA, qualsiasi scelta di politica criminale dovrebbe costituire lo *step* finale di un percorso più ampio e articolato di regolamentazione pubblica, in seno al quale il legislatore dovrebbe farsi carico anche di fissare le "regole del gioco" e i confini del c.d. *erlaubtes Risiko* ("rischio consentito")[48].

Le istituzioni dell'UE, da ultimo, hanno intrapreso azioni fondamentali nel senso auspicato, soprattutto con la proposta di regolamento del Parlamento europeo e del Consiglio COM(2021)206 – la c.d. legge sull'intelligenza artificiale[49] – recentemente approvata e di cui si attende la pubblicazione nella Gazzetta Ufficiale dell'Unione. Analoghe azioni regolatorie, su una più ampia scala globale, potrebbero essere intraprese dalle Organizzazioni internazionali, mediante la stipula di accordi multilaterali e coordinandone l'applicazione e l'attuazione[50].

Lo strumento europeo di cui si è appena concluso l'*iter* formativo mira a introdurre regole comuni nel mercato unico per garantire la circolazione di strumenti di IA sicuri, vietando talune pratiche particolarmente pericolose – tra cui alcune forme, molto controverse, di polizia predittiva: *predictive policing*[51] – e prevedendo specifiche misure di *compliance*, tra le quali il monitoraggio successivo all'immissione sul mercato del prodotto. A queste prescrizioni dovranno conformarsi in modo cogente gli operatori che vorranno

---

[48] Sull'importanza di tale aspetto, v. anche FIORELLA, *Responsabilità penale del Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in *Il diritto nell'era digitale*, cit., 656 ss.

[49] Il 21 maggio 2024 il Consiglio dell'UE ha licenziato il testo definitivo del regolamento, già oggetto di approvazione da parte del Parlamento il 13 marzo 2024, il quale apporta modifiche significative rispetto alle precedenti versioni, alcune delle quali – come vedremo – riguardano proprio l'inquadramento della *predictive policing*.

[50] Come auspicato anche dal Santo Padre Francesco nel Messaggio per la 57ma Giornata Mondiale della Pace, cit.

[51] La versione definitiva dell'*AI Act* colloca tra le pratiche vietate gli strumenti di polizia predittiva che identificano i potenziali criminali unicamente sulla base della profilazione, specificando altresì che il divieto viene meno allorché gli stessi siano utilizzati a supporto di una valutazione umana circa il coinvolgimento di un individuo in un'attività criminale che già si fondi su un ragionevole sospetto (v. art. 5, par. 1, lett. d). Tale disposizione di compromesso ha temperato il precedente, generale divieto che era stato proposto, rispetto a questi sistemi, da parte del Parlamento europeo nel giugno 2023. Cfr., per un'ampia e accurata analisi, PIETROCARLO, Predictive policing: *criticità e prospettive dei sistemi di identificazione dei potenziali criminali*", in *Dir. pen. cont.*, 2023, 2, 145 ss. Nella letteratura straniera sul tema, v., per tutti, FERGUSON, *Policing Predictive Policing*, in 94 *Washington Law Review*, 2017, 5, 1109 ss.

commercializzare *software* di IA ad alto rischio, e altrettanti obblighi graveranno sugli operatori per la distribuzione di vari dispositivi. Lo strumento normativo eurounitario prevede inoltre che gli Stati membri (SM) istituiscano o designino autorità nazionali responsabili di garantire l'applicazione di tali norme e che stabiliscano (cfr. art. 99) «le regole relative alle sanzioni» – le quali devono essere «efficaci, proporzionate e dissuasive» – «e alle altre misure di esecuzione, che possono includere anche avvertimenti e misure non pecuniarie, applicabili in caso di violazione del presente regolamento da parte degli operatori». Si specifica infine che gli SM sono tenuti ad «adottare tutte le misure necessarie» al fine di garantire un'attuazione corretta ed efficace del regolamento medesimo. Di conseguenza, il testo lascia un ampio margine di apprezzamento e decisione ai singoli Stati membri sui tipi di misure punitive da adottare (nonché sul loro contenuto), al di là dei vincoli essenziali sanciti dalla regolamentazione[52].

[52] In effetti, l'art. 99 del regolamento stabilisce anche quanto segue: «[...] 3. La non conformità al divieto delle pratiche di IA di cui all'articolo 5 è soggetta a sanzioni amministrative pecuniarie fino a 35.000.000 EUR o, se l'autore del reato è un'impresa, fino al 7% del fatturato mondiale totale annuo dell'esercizio precedente, se superiore. 4. La non conformità di un sistema di IA a qualsiasi delle seguenti disposizioni connesse a operatori o organismi notificati, diverse da quelle di cui all'articolo 5, è soggetta a sanzioni amministrative pecuniarie fino a 15.000.000 EUR o, se l'autore del reato è un'impresa, fino al 3% del fatturato mondiale totale annuo dell'esercizio precedente, se superiore: a) gli obblighi dei fornitori a norma dell'articolo 16; b) gli obblighi dei rappresentati autorizzati a norma dell'articolo 22; c) gli obblighi degli importatori a norma dell'articolo 23; d) gli obblighi dei distributori a norma dell'articolo 24; e) gli obblighi dei *deployer* a norma dell'articolo 26; f) i requisiti e gli obblighi degli organismi notificati a norma dell'articolo 31, dell'articolo 33, paragrafi 1, 3 e 4, o dell'articolo 34; g) gli obblighi di trasparenza per i fornitori e gli utenti a norma dell'articolo 50. 5. La fornitura di informazioni inesatte, incomplete o fuorvianti agli organismi notificati o alle autorità nazionali competenti per dare seguito a una richiesta è soggetta a sanzioni amministrative pecuniarie fino a 7.500.000 EUR o, se l'autore del reato è un'impresa, fino all'1% del fatturato mondiale totale annuo dell'esercizio precedente, se superiore. 6. Nel caso delle PMI, comprese le start-up, ciascuna sanzione pecuniaria di cui al presente articolo è pari al massimo alle percentuali o all'importo di cui ai paragrafi 3, 4 e 5, se inferiore. 7. Nel decidere se infliggere una sanzione amministrativa pecuniaria e nel determinarne l'importo in ogni singolo caso, si tiene conto di tutte le circostanze pertinenti della situazione specifica e, se del caso, si tiene in considerazione quanto segue: a) la natura, la gravità e la durata della violazione e delle sue conseguenze, tenendo in considerazione la finalità del sistema di IA, nonché, ove opportuno, il numero di persone interessate e il livello del danno da esse subito; b) se altre autorità di vigilanza del mercato di uno o più Stati membri hanno già applicato sanzioni amministrative pecuniarie allo stesso operatore per la stessa violazione; c) se altre autorità hanno già applicato sanzioni amministrative pecuniarie allo stesso operatore per violazioni di altre normative dell'Unione o nazionali, qualora tali violazioni derivino dalla stessa attività o omissione che costituisce una violazione pertinente del presente regolamento; d) le dimensioni, il fatturato annuo e la quota di mercato dell'operatore che ha commesso la violazione; e) eventuali altri fattori aggravanti o attenuanti applicabili alle circostanze del caso, ad esempio i benefici

Dal nostro punto di vista, prima di ipotizzare qualsiasi misura punitiva diretta contro le aziende produttrici o utilizzatrici di strumenti di IA, gli Stati – anche al di fuori dell'UE e quindi non soggetti alle venture disposizioni europee direttamente applicabili – dovrebbero adottare una legislazione che disciplini la produzione e la vendita di tali congegni tecnologici e le altre attività nel settore. Tale legislazione, come minimo, dovrebbe identificare quali prodotti sono consentiti, stabilire le regole di *compliance* da osservare per introdurre in sicurezza questi strumenti sul mercato, e imporre forme di monitoraggio post-vendita, indicando, altresì, l'autorità pubblica deputata ai controlli.

Il secondo passo da intraprendere consiste in un'attenta riflessione sul ruolo che il rimedio penalistico può assolvere in questa materia, con particolare riguardo alle azioni di *enforcement* concernenti le imprese in relazione a reati legati all'utilizzo dell'IA.

In linea teorica, potrebbero essere prese in considerazione e sviluppate tre strategie sanzionatorie.

1) La prima strategia consiste nella previsione di *sanzioni penali/punitive applicabili al sistema di IA in quanto tale*, vale a dire quale autore diretto di un reato, ovviamente previo riconoscimento della personalità giuridica. Sotto questo profilo, andrebbe considerata anche la possibilità di far scattare, in aggiunta, le ordinarie regole sulla responsabilità penale o *ex crimine* degli enti vigenti nelle varie giurisdizioni, nel caso in cui gli illeciti penali collegati alla IA siano compiuti nel contesto di un'organizzazione societaria.

2) Il secondo modello di intervento potrebbe declinarsi nell'introduzione di *sanzioni penali/punitive dirette contro l'impresa che abbia prodotto o utilizza-*

---

finanziari conseguiti o le perdite evitate, direttamente o indirettamente, quale conseguenza della violazione; f) il grado di cooperazione con le autorità nazionali competenti al fine di porre rimedio alla violazione e attenuarne i possibili effetti negativi; g) il grado di responsabilità dell'operatore tenendo conto delle misure tecniche e organizzative attuate; h) il modo in cui le autorità nazionali competenti sono venute a conoscenza della violazione, in particolare se e in che misura è stata notificata dall'operatore; i) il carattere doloso o colposo della violazione; j) l'eventuale azione intrapresa dall'operatore per attenuare il danno subito dalle persone interessate. 8. Ciascuno Stato membro può prevedere regole che dispongano in quale misura possono essere inflitte sanzioni amministrative pecuniarie ad autorità pubbliche e organismi pubblici istituiti in tale Stato membro. [...]». Su questo tema v. anche MINELLI, *La responsabilità "penale" tra persona fisica e* corporation *alla luce della Proposta di Regolamento sull'Intelligenza Artificiale,* in *Dir. pen. cont.,* 2022, 2, 50 ss.; GIANNINI, *Intelligenza artificiale,* human oversight *e responsabilità penale: prove d'impatto a livello europeo*, in *Criminalia*, 2021, 249 ss.

*to un sistema di IA*, causando danni a terzi e/o realizzando gli elementi costitutivi di un reato. Secondo questo meccanismo ascrittivo, la responsabilità autonoma e diretta dell'ente imprenditoriale potrebbe basarsi sulla mancata adozione, prima dell'evento avverso dovuto al cattivo funzionamento dell'IA, di misure di *compliance* legale con funzione preventiva. Si tratterebbe, così, di un caso di responsabilità societaria fondata su un requisito di colpa d'organizzazione.

3) La terza soluzione consiste nel comminare *sanzioni punitive alle imprese*, a prescindere dal verificarsi di danni e/o reati, in ragione della mera *violazione* dei requisiti e *degli obblighi di compliance* che devono essere soddisfatti in relazione all'introduzione di strumenti di IA sul mercato e al loro monitoraggio post-distribuzione e vendita. Questa strategia presuppone, come già segnalato, la previa adozione di una regolamentazione pubblica che disciplini olisticamente il fenomeno in esame.

Ciò detto, i tre modelli non sembrano parimenti plausibili.

Circa la *prima ipotesi* regolatoria (responsabilità diretta degli algoritmi), abbiamo già esternato le ragioni per le quali l'idea secondo cui *machina delinquere potest* non sia né convincente né praticabile. Come chiarito, lo stato attuale delle conoscenze scientifiche impedisce di cogliere nel funzionamento degli strumenti di IA una *free will* o una vera e propria autocoscienza e capacità di essere motivati dalla minaccia di conseguenze negative[53]. Legittimare una responsabilità di questo tenore sarebbe una *fictio* priva di senso, ferma la possibilità di "aggredire" materialmente lo strumento di IA in funzione dell'obiettiva pericolosità acclarata.

[53] Sul punto, v., nella letteratura italiana: BASILE, *Intelligenze artificiali e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, 2019, 2 ss.; CAPPELLINI, Machina delinquere non potest.? *Brevi appunti su intelligenza artificiale e responsabilità penale*, in *www.discrimen.it*, 27 marzo 2019; BORSARI, *Intelligenza artificiale e responsabilità penale: prime considerazioni*, in *Medialaws*, 2019, 3, 262 ss.; PIVA, *Machina discere, (deinde) delinquere et punire potest*, in *Il diritto nell'era digitale*, cit., 681 ss.; SEVERINO, *Intelligenza artificiale e diritto penale*, in *Intelligenza artificiale: il diritto, i diritti, l'etica*, a cura di Ruffolo, Milano, 2020, 533 ss.; TRIPODI, *Uomo,* societas*,* machina, in *Leg. pen.*, 2023, 12. Nella letteratura internazionale v., oltre agli autori già citati nelle note precedenti, LAGIOIA-SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in 33 *Philosophy & Technology*, 2020, 433 ss.; LIMA, *Could AI Agents Be Hold Criminally Liable?: Artificial Intelligence and The Challenges for Criminal Law*, in 69 *South Carolina Law Review*, 2018, 677 ss.

La *seconda soluzione* è, in linea di principio, percorribile[54]. Nondimeno anch'essa sollecita alcune puntualizzazioni e qualche *caveat*. Introdurre una disposizione punitiva per l'omessa prevenzione del reato-evento determinato dal malfunzionamento del sistema di IA, in assenza di una regolamentazione pubblica che detti, in modo puntuale, gli *standard* di produzione e vendita di tali strumenti tecnologici, equivarrebbe ad affidare – in modo irragionevole e dunque ingiusto – alle imprese il compito di gestire tutti i rischi (e le correlate questioni di responsabilità) connessi alla produzione e all'impiego di questi congegni.

Pertanto, come abbiamo già rimarcato in precedenza, in ossequio a un principio generale di politica del diritto razionale, l'introduzione di sanzioni – formalmente o sostanzialmente – penali contro le società dovrebbe costituire l'approdo finale di una più vasta e articolata strategia di regolamentazione pubblica. Inoltre, perlomeno in taluni scenari nient'affatto remoti, è discutibile che possa legittimarsi – sul piano dommatico e politico-criminale – una disposizione che introduca la responsabilità penale delle società anche per la causazione materiale di esiti negativi imprevedibili, generati da sistemi capaci di decisione completamente autonoma. In altre parole, si può dubitare fortemente della possibilità di dimostrare in giudizio la colpa d'organizzazione di un ente, se nemmeno con tutta la diligenza astrattamente pensabile l'azienda avrebbe potuto prevedere e dunque prevenire il danno causato dall'"inceppamento" di un sistema di IA il cui funzionamento risulti – del tutto o almeno in parte – insondabile[55]. È evidente che, in questo caso, l'unica reale possibilità di evitare l'evento imprevedibile sarebbe rinunciare *in toto* alla produzione e/o all'impiego del sistema di IA, piegandosi a una regola di astensione a base precauzionale.

Tenendo conto dell'insieme di questi aspetti, la *terza soluzione* potrebbe

[54] Cfr. anche PANATTONI, *AI and Criminal Law*, cit., 125 e ss. Per una panoramica generale, cfr. GRUODYTE-CERKA, *Artificial Intelligence as a Subject of Criminal Law: A Corporate Liability Model Perspective*, in *Smart Technologies and Fundamental Rights,* a cura di Gordon, Leiden, 2020, 260 ss. Per quanto riguarda alcune di queste prospettive future, v. anche *Feasibility study on a future Council of Europe instrument on artificial intelligence and criminal law*, pubblicato dal Comitato europeo sui problemi della criminalità (CDPC) del Consiglio d'Europa nel 2020, disponibile su *www.coe.int/cdpc*.

[55] Sulla valutazione della colpa organizzativa d'impresa nel contesto della nuova era digitale v., altresì, NISCO, *Riflessi della* compliance *digitale in ambito 231*, in *www.sistemapenale.it*, 14 marzo 2022.

rappresentare, quantomeno nelle fasi genetiche del processo regolatorio che la materia reclama, un compromesso equilibrato per evitare, da un lato, ingiuste imputazioni di responsabilità e, dall'altro, una sostanziale paralisi della ricerca e della produzione nel campo della IA, con un indesiderabile effetto *chilling*, cioè di "congelamento". Come abbiamo già sottolineato, per sviluppare al meglio questa politica, il primo passo deve essere l'introduzione d'una legislazione che, in linea con il nuovo regolamento europeo citato, detti le "regole del gioco" e tutti i requisiti di conformità connessi all'immissione sul mercato e al monitoraggio degli strumenti di IA. Questa opzione appare anche più rispettosa del principio di *ultima ratio*, secondo cui il diritto penale dovrebbe essere attivato solo dopo aver verificato la capacità di altre forme di responsabilità – *in primis*, la responsabilità civile e quella amministrativa – di assicurare una tutela adeguata agli interessi in gioco, tenuto conto anche della novità del fenomeno da gestire e dei rischi da affrontare. A questo riguardo, merita un cenno anche la risoluzione del Parlamento UE del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, la quale aveva suggerito, tra l'altro, l'istituzione di un "regime assicurativo obbligatorio" per i potenziali danni correlati a tali contesti, nonché un «fondo per garantire la possibilità di risarcire i danni in caso di assenza di copertura assicurativa»[56].

Pertanto, solo dopo aver disegnato un quadro regolatorio esaustivo, in grado di governare in modo accorto il fenomeno esaminato, diverrebbe ragionevole la previsione di sanzioni dirette all'ente collettivo, tali da spostare il *focus* delle attività di *enforcement* riguardanti le persone giuridiche dalla causazione di danni a terzi[57] alla violazione degli obblighi di *compliance* per l'introduzione sicura nel mercato e il monitoraggio post-vendita di strumenti di IA. Tali congegni sanzionatori, infatti, mirerebbero a garantire la corretta applicazione delle normative pubbliche da parte degli operatori economici e si porrebbero

[56] Il testo della risoluzione è disponibile al seguente link: *https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html*. Su questo tema v. anche la Proposta di direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), COM/2022/496 final.

[57] Questo settore, come già detto, potrebbe essere coperto almeno inizialmente da strumenti di responsabilità civile.

a presidio del rispetto di tali (essenziali) previsioni. La regolamentazione pubblicistica di settore potrebbe sancire, in applicazione del principio di precauzione, anche divieti assoluti di produzione e commercializzazione di specifiche tecnologie che non diano minime garanzie di affidabilità e che siano ritenute troppo rischiose per i danni che potrebbero ingenerare.
In questo contesto, sarebbe auspicabile una strategia sanzionatoria articolata, composta da misure punitive classiche e misure "programmatiche" o di taglio ingiunzionale. Dal primo punto di vista, andrebbero introdotte sanzioni pecuniarie per la violazione delle regole di produzione/commercializzazione/utilizzo dei prodotti di IA o per la violazione del dovere di segnalazione alle autorità in caso di incidenti dannosi o pericolosi. Dal secondo versante, dovrebbe riconoscersi al giudice il potere di ricorrere a ingiunzioni per imporre alle aziende l'adempimento doveroso e segnatamente l'innesto o il miglioramento dei presidi di *compliance* aziendale e dei sistemi di controllo interno[58], a cui potrebbe affiancarsi un periodo di monitoraggio pubblico (amministrativo o giudiziario, comunque da parte di una istituzione pubblica), per verificare scrupolosamente il reale adeguamento agli *standard* imposti dalla disciplina di settore[59].
Inoltre, perlomeno in una prima fase di "sperimentazione" della nuova legislazione, è auspicabile un *enforcement* puramente amministrativo ai fini dell'eventuale applicazione di queste sanzioni basate sul mancato rispetto della precipua regolamentazione in tema di IA. Ciò consentirebbe, nel corso del tempo, anche di comprendere se un modello di disciplina basato su disposizioni non penali sia sufficiente a contrastare danni potenzialmente generabili da dispositivi di IA.

4. *Conclusioni*. In questo contributo abbiamo cercato di scandagliare le opportunità e le sfide che deve affrontare l'eventuale responsabilizzazione delle

[58] In senso analogo, v. CONSULICH, Flash offenders. *Le prospettive di* accountability *penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.*, 2022, 3, 1051 ss.
[59] Sulle possibili riforme del sistema sanzionatorio a carico degli enti collettivi nell'ordinamento italiano, sia consentito rinviare, anche per una più ampia rassegna della letteratura, a MONGILLO, *Il sistema delle sanzioni applicabili all'ente collettivo tra prevenzione e riparazione. Prospettive* de iure condendo, in *Riv. trim. dir. pen. ec.*, 2022, 3-4, 559 ss.

società in relazione a reati legati all'utilizzo di sistemi di IA.
Con riferimento ai dispositivi completamente autonomi, si è evidenziato come allo stato emerga un vuoto di responsabilità, considerati i modelli di *corporate criminal liability* diffusi a livello internazionale.
Si è anche messo in risalto come la previsione di sanzioni penali rivolte direttamente contro i sistemi digitali di IA e la stessa punizione "diretta" degli enti che ne facciano uso non trovino un basamento adeguato nell'attuale stadio delle conoscenze scientifiche.
In questo contesto, le uniche alternative plausibili sembrano ridursi all'innesto di misure punitive per colpire: *(a)* l'organizzazione che non metta in atto, negligentemente, misure di controllo atte a contenere il rischio di verificazione di risultati avversi dovuti all'IA, nei casi e nella misura in cui le conoscenze tecnico-scientifiche oggi disponibili permettano l'individuazione di misure di prevenzione di un rischio calcolabile: in questo caso l'ente risponderebbe secondo il modello dell'omessa prevenzione dei danni verificatisi a causa dell'utilizzo del dispositivo di IA; ovvero *(b)* la *societas* che ometta di adottare misure di *compliance* legale o gli *standard* prestabiliti per produrre e commercializzare legalmente tali prodotti digitali, a prescindere dalla realizzazione di un reato o risultato lesivo.
Abbiamo, quindi, cercato di segnalare i punti di forza e di debolezza di entrambe queste opzioni, la cui messa a fuoco rivela quanto sia difficile regolamentare un universo tecnologico che sta crescendo a velocità inimmaginabili fino a pochi anni fa, e quali complesse implicazioni debbano essere considerate nello sviluppo delle necessarie azioni di politica legislativa[60].
Nell'affrontare queste sfide, le strategie di efficace contenimento dei rischi insiti nell'utilizzo delle nuove forme di IA devono puntare allo sviluppo di tecnologie al servizio della dignità umana, dei diritti fondamentali e del progresso sociale, senza sacrificare – nella materia sui cui ci siamo concentrati – i principi fondamentali del diritto penale, quali canoni inderogabili di civiltà giuridica.

[60] V. anche BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019, 4, 1909; LAUFER, *The Missing Account of Progressive Corporate Criminal Law*, in 14 *New York University Journal of Law & Business*, 2017, 71.