amministrativ@mente
Rivista scientifica trimestrale di diritto amministrativo
www.amministrativamente.com

UNIVERSITÀ DEGLI STUDI DI ROMA "FORO ITALICO"

Rivista scientifica trimestrale di diritto amministrativo (Classe A)

Pubblicata in internet all'indirizzo www.amministrativamente.com

Rivista di Ateneo dell'Università degli Studi di Roma "Foro Italico"

Direzione scientifica
Gennaro Terracciano, Gabriella Mazzei, Julián Espartero Casado

Direttore Responsabile
Gaetano Caputi

Redazione
Carlo Rizzo

FASCICOLO N. 3/2025

Estratto

Iscritta nel registro della stampa del Tribunale di Roma al n. 16/2009

ISSN 2036-7821

Comitato scientifico

Annamaria Angiuli, Antonio Barone, Vincenzo Caputi Jambrenghi, Francesco Cardarelli, Enrico Carloni, Maria Cristina Cavallaro, Guido Clemente di San Luca, Andry Matilla Correa, Chiara Cudia, Gianfranco D'Alessio, Mariaconcetta D'Arienzo, Ambrogio De Siano, Ruggiero Dipace, Valentina Maria Donini, Luigi Ferrara, Pierpaolo Forte, Gianluca Gardini, Biagio Giliberti, Emanuele Isidori, Bruno Mercurio, Francesco Merloni, Giuseppe Palma, Alberto Palomar Olmeda, Attilio Parisi, Luca Raffaello Perfetti, Fabio Pigozzi, Alessandra Pioggia, Helene Puliat, Francesco Rota, Josè Manuel Ruano de la Fuente, Leonardo J. Sańchez-Mesa Martĭnez, Ramón Terol Gómez, Antonio Felice Uricchio.

Comitato editoriale

Jesús Avezuela Cárcel, Giuseppe Bettoni, Sveva Bocchini, Salvatore Bonfiglio, Vinicio Brigante, Sonia Caldarelli, Giovanni Cocozza, Andrea Marco Colarusso, Sergio Contessa, Beatrice Coppa, Giuseppe Doria, Manuel Delgado Iribarren, Fortunato Gambardella, Flavio Genghi, Jakub Handrlica, Margherita Interlandi, Laura Letizia, Giuseppina Lofaro, Federica Lombardi, Gaetano Natullo, Carmen Pérez González, Giovanni Pesce, Benedetta Piazza, Marcin Princ, Sara Pugliese, Bianca Nicla Romano, Antonio Saporito, Giuliano Taglianetti, Esper Tedeschi, Simona Terracciano, Stefania Terracciano, Salvatore Villani.

Coordinamento del Comitato editoriale

Valerio Sarcone.

ISSN 2036-7821

amministrativ@mente
Rivista scientifica trimestrale di diritto amministrativo
www.amministrativamente.com

UNIVERSITÀ DEGLI STUDI DI ROMA "FORO ITALICO"

Il supporto dell'intelligenza artificiale nell'elaborazione dei "*compliance programs* 231" e i problematici profili della responsabilità penale dei programmatori e dei sorveglianti dei *software*

di Francesco Marzullo

(Abilitato alle funzioni di professore associato di diritto penale)

Sommario

1. Il punto: il ricorso alla I.A., nel sistema della responsabilità da reato degli enti. – 2. Legittimazione e ruolo dell'I.A. nell'elaborazione del M.O.G.: il problematico accertamento della responsabilità amministrativa dell'ente. - 3. La tesi di Piergallini circa la responsabilità diretta della macchina artificiale. – 4. Soluzioni *de lege ferenda* per promuovere l'impiego dell'I.A. nella costruzione dei M.O.G. – 5. L'inedita dimensione antropica dell'imputazione dei "reati presupposto 231": la responsabilità penale dei programmatori/produttori dei *software* utilizzati nella redazione dei protocolli cautelari. – 5.1. - La responsabilità penale dei c.d. "sorveglianti". - 6. Conclusioni.

Abstract

Due to the constant circularity and widespread diffusion of the culture of compliance and compliance programs, typified by Legislative Decree 231/2001 within complex corporate organizations (whose preparation and implementation results in the legal person not being punishable under corporate criminal liability), the Author addresses an absolutely open and multidisciplinary topic, which currently registers limited interest from the most accredited doctrine: that is, the usefulness of using, in the immediate future, technologies emerging from Artificial Intelligence, to develop precautionary protocols capable of not only attenuating, but even neutralizing any risk of verification of the underlying crime (according to the paradigm of attribution of criminal liability through negligence). The analysis then focuses, in an anthropic dimension of the accusation, on the problem of the possible criminal liability of human programmers and the so-called "supervisors" of those software which, although used in the drafting of precautionary protocols, they have not had any effectiveness in preventing the risk of crime.

* *Il presente lavoro è stato sottoposto al preventivo referaggio secondo i parametri della double blinde peer review*

ISSN 2036-7821

1. Il punto: il ricorso alla I.A., nel sistema della responsabilità da reato degli enti.

Nel grande cantiere sempre aperto della responsabilità da reato degli enti, una nuova e complessa area è rappresentata dalla emersione di una casistica allargata, relativa al possibile ricorso alla I.A. nella predisposizione prima e nell'attuazione poi dei modelli organizzativi previsti in materia di *corporate criminal liability.*

Si tratta di una tematica assai complessa, aperta alla integrazione di competenze plurispecialistiche e non esclusivamente giuridiche e, preventivamente, condizionata dalla mancanza di una definizione unitaria di I.A.[1].

Per una necessaria, quanto sintetica, *actio finium regundorum*, occorre concettualmente distinguere tra IA forti e IA deboli[2] : le prime astrattamente deputate alla contestualizzazione delle istanze di analisi; mentre alle seconde resta assegnata la fase esecutiva degli ambiti di ricerca [3].

Ed infatti, mentre nella versione "forte", l'IA punta a conseguire un livello di intelligenza pari a quello dell'uomo (*human-like)*; nella prospettiva "debole", l'operatività della macchina si arresta ad una funzione meramente sussidiaria [4], che sul terreno operativo apre a tre passaggi essenziali: la percezione dei dati dall'ambiente in cui il sistema è immerso; l'elaborazione delle informazioni da questi estrapolate, attraverso cui individuare l'azione migliore per il raggiungimento del risultato prefissato e l'attuazione della c.d. "azione decisa".

La sequenza dei citati passaggi consente il raggiungimento dell'obiettivo finale rappresentato dall'algoritmo, come « schema esecutivo della macchina che memorizza tutte le opzioni decisionali in base ai dati che progressivamente elabora».

In sintesi estrema, gli elementi alla base del funzionamento di ogni sistema di I.A. sono: i dati raccolti ed immessi nel sistema – dalla cui qualità dipende la qualità del sistema e quindi dell'output offerto – l'algoritmo elaborato dal programmatore, che consente di ottenere un certo risultato partendo dalle informazioni d'ingresso.

Con l'efficace espressione "*compliance digitale*"[5] si è inteso, dunque, evocare gli effetti dell'applicazione di tecnologie emergenti basate sull'Intelligenza Artificiale (I.A.) ai sistemi di *compliance* e, tra questi, ai modelli di organizzazione e gestione previsti dal

---

[1] C. PIERGALLINI, *Intelligenza artificiale: da "mezzo" a "autore" del reato?,* in *Riv. it. dir. proc. pen.*, 2020, p. 1745 ss.: G. BARONE, *Intelligenza artificiale e processo penale: la linea dura del Parlamento europeo. Considerazioni a margine della risoluzione del Parlamento europeo del 6 ottobre 2021*, in *Cass. pen.*, 2022, 3, pp. 1180 – 1191.

[2] C. PIERGALLINI, *Intelligenza artificiale: da "mezzo" a "autore" del reato?,* op. cit., p. 1746 ss.

[3] Secondo la C.E.P.E.J., *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, 2018, p. 47, nel prossimo futuro, difficilmente potranno esistere sistemi di IA forte, posto che le IA per essere in grado di modellizzare il mondo nella sua interezza, necessiterebbero di progressi significativi della ricerca di base e non soltanto di semplici miglioramenti delle prestazioni dei sistemi esistenti.

[4] Si veda *AI-HLEG, High-Level Expert Group on Artificial Intelligence, A definition of Artificial Intelligence: maincapabilities and scientificdisciplines*, Bruxelles, 18 dicembre 2018, consultabile sul sito della Commissione europea; *JRC Technical Reports, AI Watch. Defining Artificial Intelligence, Publications Office of the European Union*, Lussemburgo, 27 febbraio 2020.

[5] L'espressione A. NISCO, *Riflessi della compliance digitale in ambito 231*, in *Sistema penale.it*, 14 marzo 2022.

ISSN 2036-7821

d.lgs. 231/2001, nonché ai protocolli cautelari che sono la declinazione dettagliata dell'autonormazione privata[6]. Tale argomento rientra nel più generale tema dell'IA applicata ai *Big Data*[7]

La piena comprensione del processo decisionale di un'I.A. rappresenta forse un risultato impossibile: il tasso di opacità tecnologica dei tipi più evoluti di intelligenza artificiale, che si caratterizzano per la capacità di apprendere e reagire in modo indipendente dall'ambiente, è piuttosto elevato, in ragione sia della complessità degli algoritmi impiegati, sia della mole dei dati elaborati[8].

La rivoluzione *cybernetica*, la creazione di *software* di intelligenza artificiale e la programmazione di algoritmi specifici, da un canto hanno inciso profondamente nel tradizionale sviluppo delle attività umane, e dall'altro hanno dato forma a nuove aree criminologiche (in cui sono più forti i legami con gli istituti di matrice penalistica) che la dottrina[9] ha subito classificato in:

i) attività di *law enforcement* e di giustizia predittiva (si pensi alle macchine utilizzate dalle forze di polizia, per lo più in via sperimentale, per compiti come

[6] Su tale specifico argomento si rinvia altresì alle riflessioni di G. IMBRIGIOTTA, *Intelligenza Artificiale e MOG 231: dalla colpa d'organizzazione alla colpa "artificiale"*, in *Rivista penale Diritto e Procedura.it*, 14 gennaio 2025; di R. TREZZA, ***L'Intelligenza Artificiale come ausilio alla standardizzazione del modello 231: vantaggi "possibili" e rischi "celati"***, in ***Giurisprudenza penale***, 2021, p. 3 .

[7] C. BURCHARD, *Digital Criminal Compliance*, in *Festschrift für U. Sieber zum 70. Geburtstag, II*, Duncker & Humblot, Berlin, 2021, p. 742 ss.; nonché P.A. RAMIREZ BARBOSA, *Corporate Criminal Law, Artificial Intelligence and Big Data: the Huawei Case and its Implications for Global Society*, in *Revista Paradigma*, v. 29, n. 1, 2020, p. 19 ss.; A. ELLIOTT, *La cultura dell'intelligenza artificiale. Vita quotidiana e rivoluzione digitale* (trad. it., di *The Culture of AI. Everyday Life and the Digital Revolution*, 2019), Codice Edizioni, Torino, 2021.

[8] La capacità di elaborare *big datada* parte dell'intelligenza artificiale impone una adeguata sorveglianza sul potere che ne deriva; si tratta di elaborare una vera e propria Big Data Policing, allo stato *in fieri* solo negli Stati Uniti, si veda A.G. F ERGUSON, *The Rise of Big Data Policing. Surveillance, Race, and the Future of Law Enforcement*, New York, 2017, 2, 167 ss.. Si rinvia ai contributi relativi alla "*Round Table on Big Data and Criminal Law*" pubblicati su *Ohio State Journal Criminal Law*, 2018, 473 ss.

[9] Per una disamina più in generale dei possibili intrecci tra IA e diritto penale, v.: U. PAGALLO, *Saggio sui robot e il diritto penale,* in S. VINCIGUERRA-F. DASSANO (a cura di), *Scritti in memoria di Giuliano Marini*, Napoli, 2010, 595 ss.; J. CHARPENTIER, *Justice Machines. Racconto di fantascienza giudiziaria*, Macerata, 2015; S. RIONDATO, *Robot: talune implicazioni di diritto penale*, in P.MORO-C. SARRA, *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Milano, 2017, 85 ss.; A. GARAPON-J. LASSÉGUE, *Justice digital. Revolution graphique et ropture anthropologique,* Paris, 2018; M. B. MAGRO, *Relazione su "Biorobotica, robotica e diritto penale"*, in *dirittopubblico.unipd.it*, 2018, 1 ss.; F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, pubblicato pure su *Diritto penale contemporaneo*, 29 settembre 2019, 1 ss.. A. CAPPELLINI, *Profili penalistici delle self-driving cars*, in *Diritto penale contemporaneo*, 2019, 2, 326 ss. Sull'uso dei robot in medicina, v. A. PERIN, Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale, in *Rivista di BioDiritto*, 2019, 1, 207 ss. Sul distinto versante del ricorso all'IA, come strumento di *lawenforcement*, di *policinge* e di giustizia predittiva, v . M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 2018, 3, 872 ss.; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen*., 2019, 1909 ss.; A. GULLO, *Nuove frontiere tecnologiche e sistema penale: alcune note introduttive,* in *Diritto penale contemporaneo*, 2019, 2, XI ss.; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, in A. CADOPPI - S. CANESTRARI- A. MANNA -M. PAPA, *Trattato di diritto penale - Cybercrime*, Utet, 2019; F. SGUBBI, *Il diritto penale totale,* Bologna, 2019, pp. 40-44; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *disCrimen*, 15 maggio 2020, 1 ss.

ISSN 2036-7821

pattugliamento, sorveglianza, disinnesco di bombe, riconoscimento facciale, ecc.);
ii) impiego di algoritmi decisionali, dopo una sperimentazione per la decisione di liti civili o di contenziosi amministrativi, anche in ambito penale;
iii) uso di c.d. algoritmi predittivi (*risk assessments*) per la valutazione, in senso ampio, della pericolosità criminale ai fini dell'applicazione di una misura di sicurezza, di una misura cautelare o di una misura di prevenzione o anche per concedere la sospensione condizionale di una pena o l'affidamento in prova al servizio sociale;
iv) coinvolgimento di un sistema di *IA* come strumento, autore o vittima di un reato.
E tuttavia, accanto a questi temi, ad avviso di chi scrive, si impone la necessità di ulteriormente riflettere sulla ammissibilità del ricorso alla creazione di algoritmi in grado di raccogliere norme giuridiche ed informazioni aziendali, per valutare il *risk assestment* e predisporre -nella fase genetica del prodotto autonormato (il modello organizzativo)- le misure più idonee a prevenire la commissione di reati o la verificazione di eventi il cui rischio è stato valutato; nonché sull'efficacia eventualmente "esimente" di quei prodotti autonormati "concepiti" dall'I.A.
Premesso che la redazione del modello implica una conoscenza effettiva dell'ente e del suo organigramma, del suo oggetto sociale, delle varie funzioni e della struttura degli organi dirigenziali, e premesso che, sulla base di tali informazioni, va formulata una puntuale mappatura dei rischi di commissione di reati, si deve prendere atto, che alcune società hanno già iniziato a sperimentare *software* capaci di progettare e redigere autonomamente i modelli organizzativi e di gestione (inclusi i relativi protocolli cautelari)[10].
E quindi il nuovo *focus* riguarda la verifica dei benefici e dei rischi della *digital criminal compliance*" e, soprattutto, le implicazioni che dall'utilizzo di tali strumenti possano derivare all'accertamento della responsabilità amministrativa dell'ente.
Il tema si complica anche in ragione di un ulteriore profilo che attiene all'utilizzo o meno di *software*, dotati di *Machine Learning* (capacità di autoapprendimento); ci riferiamo a quei programmi in grado di acquisire dati dalla propria esperienza, al fine di elaborare nuove strategie per affrontare problemi futuri.
La tecnologia basata sulle reti neurali è alla base della gran parte dei motori per l'apprendimento delle IA e attraverso il c.d. processo di *machine learning*, una rete neurale viene addestrata, a riconoscere gli scenari che tali dati rappresentano.

[10] N. ABRIANI, *La corporate governance nell'era dell'algoritmo Prolegomeni a uno studio sull'impatto dell'intelligenza artificiale sulla* corporate governance, in *Il nuovo diritto delle società*, 2020, 3, p. 268 ss., laddove l'Autore individua i principali settori in cui si inizia ad intravedere un'interferenza tra intelligenza artificiale e *governance* aziendale, ed in particolare: i) nell'utilizzo di strumenti di IA per l'organizzazione e il funzionamento societario interno (prospettiva interna di corporate governance); ii) negli strumenti di IA sia come supporto sia come output dell'attività di impresa; iii) nel ricorso a strumenti di IA per la valutazione dall'esterno del funzionamento societario (prospettiva esterna dei mercati finanziari).

ISSN 2036-7821

Dopo la fase di *machine learning*, una rete neurale diventa capace di riconoscere gli scenari che ha "imparato" nei nuovi dati che riceve in input. Inoltre, una rete neurale potrebbe continuare ad "imparare" nuovi scenari dai dati ricevuti come *input*.
Si tratta, a ben vedere, di vere e proprie macchine pensanti, la cui caratteristica è quelle di essere autonome: svolgono cioè attività in assenza del controllo umano e sono in grado di apprendere dalla propria esperienza e di migliorare il comportamento[11].
L'ideatore (umano) genera l'algoritmo, che 'conosce', grazie all'esperienza, e 'rielabora', potendo pervenire a scelte del tutto impreviste ed imprevedibili da parte dell'ideatore.
La funzione innovativa di questi sistemi di intelligenza artificiale, dotati di una sorta di "autocoscienza", potrebbe consentire di individuare tempestivamente le carenze organizzative dell'ente, sulla base dei rischi rilevati in situazioni analoghe pregresse[12], al fine di dotare la persona giuridica di strumenti prevenzionistici sempre attuali e più efficaci.
E tuttavia cosa potrebbe accadere sul piano delle eventuali responsabilità degli attori collettivi ex decreto 231/2001, se anche questi modelli in astratto "più affidabili", in quanto predisposti da sistemi digitali più innovativi ed evoluti, fossero invece ritenuti inidonei in fase di validazione giudiziale? Ed ancora: sarebbe in qualche modo ipotizzabile un intervento punitivo diretto[13] nei confronti dell'agente

[11] Si tratta in particolare delle forme di intelligenza artificiale di quarto livello o sistemi multi-agente (c.d. multi-agent systems), su cui J. XIE-C.C. LIU, *Multi Agent Systems and Their Application*, in *Journal of International Council on Electrical Engineering*, 2017, 188 ss. Sulla capacità di autoapprendimento come criterio per distinguere il livello di autonomia di un agente artificiale, P. NORVIG, *Macchine che apprendono*, in D. HEAVEN (a cura di), *Macchine che*

pensano. La nuova era dell'intelligenza artificiale, Bari, 2018, p. 31 ss.

[12] I *software* di intelligenza artificiale hanno la possibilità di apprendere strategie efficaci, già adottate da altre macchine in situazioni analoghe, grazie ad un *cloud* comune, modificandole ed adattandole al caso concreto senza la necessità di un nuovo intervento del programmatore grazie al cd. *black box algorithm*. Sul punto si rinvia a F. TRIPODI, *Uomo, Societas, Machina*, in *La legislazione penale*, 2023, p. 11, laddove, per un approfondimento sul tema, si cita P. NORVIG, *Macchine che apprendono*, in D. HEAVEN (a cura di), *Macchine che pensano. La nuova era dell'intelligenza artificiale*, Bari, 2018, pp. 31 e ss.; si veda altresì S. BECK, *Intelligence agents and criminal law – Neglicence diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, 2016, p. 138 e ss.

[13] Il più noto sostenitore della teoria della responsabilità penale dell'AI è il penalista israeliano G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities, From Science Fiction to Legal Social Control*, in 4 *Akron lntel. Prop. J.*, 2010, 189, 191 ss., ID, *Liability for Crimes Involving Artificial Intelligence Systems*, Berlin, 2015, 47 ss.; tra gli altri Autori, meritano di essere citati: Y. HU, *Robot criminals*, in *52 Mich. J . L. Reform, 2019*, 487, p. 531. Nella dottrina portoghese si segnala, a titolo di esempio, P.M. FREITAS - F . ANDRADE - P. NOVAIS, *Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible*, in P. CASANOVAS - U. PAGALLO - M. PALMIRANI - G. SARTOR (a cura di), *AI Approaches to the Complexity of Legal Systems*, Berlin-Heidelberg, 2014, 145 ss.; tra gli Autori tedeschi, S. BECK, *Grundlegende Fragen zum rechtlichen Urnmgangrnit der Robotik*, in *JR, 6/2009*, p. 230 ss.; più di recente E. HILGENDORF, *Zivil-und strafrechtliche Haftung für von Maschinen verursachte Schädene*, nonché M. SIMMLER, *Maschinenetik und strafrecht-liche Verantwortlichkeit*, entrambi in O. BENDEL (a cura di), *Handbuch Maschinenetik, Wiesba-den, 2019*, 437 ss. e p. 453 ss.; scettico sulla possibilità di imputare il fatto dell'agente umano S.BECK, *Intelligent Agents and Criminal Law. Ne gligence, Diffusion of Liability and Electronic Personhood*, in *Robotics an*

ISSN 2036-7821

artificiale? Oppure la responsabilità degli eventi lesivi che derivano dall'attività degli agenti artificiali[14] potrebbe ricadere esclusivamente sugli operatori umani (l'utilizzatore, il produttore, il programmatore, dello sviluppatore, etc.) che mantengono la *signoria* sul *software*, sul processo decisionale e sul comportamento dell'algoritmo.

2. *Legittimazione e ruolo dell'I.A. nell'elaborazione del M.O.G.: il problematico accertamento della responsabilità amministrativa dell'ente.*

È un dato ormai pacifico, che l'intelligenza artificiale sia -nel contempo- un validissimo strumento da integrare nella *compliance* al servizio dei modelli comportamentali degli amministratori, per ridurre al minimo i fattori di rischio d'impresa, nonché al servizio degli organi preposti alla vigilanza per consentire un intervento immediato in presenza di segnali di allarme chiaramente percepibili

E si tratta, al tempo stesso di un tema del quale la dottrina[15] ha preso già consapevolezza, classificando con la locuzione *Digital Criminal Compliance* l'intera tipologia dei servizi di gestione digitale della *compliance* presenti sul mercato.

Come è stato efficacemente rilevato[16], se gli algoritmi forniscono i dati rilevanti in "tempo reale", in una griglia temporale predefinita, le c.d. analisi in tempo reale, permettono di seguire le strutture comportamentali nel loro processo di sviluppo e di distinguere qualitativamente se e quando le regole saranno probabilmente ignorate o alla fine rispettate.

L'intelligenza artificiale potrebbe pertanto essere messa al servizio del *management* in

---

Autonomous Systems, 2016, p. 138 ss.. Per una panoramica sul dibattito in materia vd. A. GIANNINI, *Criminal behavior and Accountability of Artificial Intelligence Systems*, Eleven, 2023.

[14] Sulla possibilità di imputare il fatto all'agente artificiale F. LA GIOIA-G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *33 Philosophy & Technology*, 2010, pp. 458 ss., che in particolare rilevano come perché una data entità sia considerata persona per il diritto penale in fondo occorre semplicemente che possano essere posti in capo ad essa, da una norma di legge, dei doveri di condotta la cui violazione implichi una sanzione. Si tratta di un'opinione non isolata se si guarda alle riflessioni di coloro che pensano possibile rimproverare per dolo o colpa gli agenti artificiali, in quanto capaci di agire, di conoscere e volere e dunque di dolo e colpa, nonché di suscettibilità alla pena in termini di riabilitazione e incapacitazione (210 ss.); cfr. ad esempio, G. HALLEVY, *The Criminal Liability of Artificial lntelligence Entities. From Science Fiction to Legal Social Control*, in 4 *Akron lntel. Prop. J.*, 2010, 186 ss.; *Y. HU, Robot Criminal Liability Revisited*, in J. SOO YOON-S. HOON HAN, S. JOAHN (a cura di), *Dangerous Ideas in Law*, Bobmunsa, 2018, 494 ss. Il presupposto sarebbe l'attribuzione a tali entità di puro pensiero elettronico di una personalità giuridica sui generis, di carattere informatico (c.d. *e-personhood*). Precursore è stato L.B. SOLUM, *Legal Personhood for Artificial Intelligence*, in 70 *N.C.L. Rev.*, 1992, p. 1231 ss.; poi ripreso da M. SIMMLER-N. MARKWALDER, *GuiltyRobots? - Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Criminal L. Forum*, 2019, 1 s. Un quadro del problema si rinviene in K. GAEDE, *Künstliche Intelligenz - Recht und Strafen für Roboter, Baden-Baden*, 2019, 31 ss.

[15] A NISCO, op. cit., p. 4 ss

[16] S. NEUFANG, *Digital Compliance – Wie digitale Technologien Compliance-Verstöße vorhersehen, in IRZ - Zeitschrift für Internationale Rechnungslegung*, 2017, p. 249. Per una trattazione esemplificativa delle modalità operative della compliance digitale, si rinvia A. SCHEMMEL, *"Effective Corporate Governance" by Legal Tech/Digital Compliance*, in S. BREIDENBACH-F. GLATZ (Hrsg.), *Handbuch Legal Tech*, 2. *Aufl.*, C.H. Beck, München, 2021, p. 166 ss.

ISSN 2036-7821

generale (organi apicali e operativi all'interno delle strutture societarie complesse), quale supporto nell'attuazione dei compiti organizzativi ed in particolare nella formazione dei processi decisionali[17] all'interno degli organi di vertice (in particolare i C.d.A.).

Con particolare riguardo all'obbligo di dotare la società di assetti organizzativi adeguati (art. 2086 c.c.), non si può escludere, che l'utilizzo dell'algoritmo possa essere impiegato alla stregua di uno strumento tecnologico irrinunciabile per poter individuare tempestivamente la ricorrenza di quei segnali di allarme, che impongono l'obbligo di attivazione in capo ai vertici aziendali (i garanti).

Il *software* potrebbe non solo favorire la percezione di segnali di allarme (la sopravvenienza di indici predittivi di crisi o di insolvenza, lo squilibrio dei flussi di cassa, le ricadute "erosive" del patrimonio netto di talune operazioni commerciali), ma soprattutto potrebbe suggerire azioni impeditive capaci di intervenire in tempo reale per neutralizzare o mitigare le fonti di rischio.

Sempre in dottrina[18], ci si è interrogati sull'eventuale ricorso ad un *software* quale "sostituto" dell'Organismo di vigilanza previsto dall'art. 6 del decreto 231.

La risposta è stata negativa in ragion dell'ineludibile componente umana (che non si può sopprimere) nell'ottemperanza dei compiti di vigilanza sull'attuazione dei protocolli cautelari; pur tuttavia non si può escludere che l'Organismo controllo possa, o debba, dotarsi di tecnologie informatiche[19] in grado di individuare rischi di non conformità del MOG, nel modo più rapido ed efficace.

Siamo così arrivati al punto nevralgico della nostra disamina: le nuove tecnologie possono o potranno essere utilizzate quale strumento per predisporre e migliorare la *compliance*, ed in particolare potranno supportare l'attore collettivo nella predisposizione dei M.O.G.?[20]

E, ancora, potrebbero residuare profili di responsabilità *ex delicto* a carico dell'ente nei casi in cui il Tribunale dovesse formulare, in sede di validazione giudiziale, un giudizio di *inidoneità* di quel modello *predisposto dall'intelligenza artificiale* e adottato dalla società?

---

[17] R.M. AGOSTINO, *Intelligenza artificiale e processi decisionali. La responsabilità degli amministratori di società,* in *Mercato concorrenza regole*, 2020, 2, p. 371 ss.

[18] A. NISCO, op.cit. p. 9 laddove l'Autore conia un'altra efficace espressione, parlando di *OdV algoritmico.*

[19] Tale proposta è stata avanzata da Una soluzione simile è stata prospettata da S. PREZIOSI, *Responsabilità da reato degli enti e intelligenza artificiale*, in *Resp. amm. soc. enti*, 2020, 4, p. 173 ss.,

[20] Un tema affine è stato recentemente trattato da D. COSTA, in *L'impatto dell'Intelligenza Artificiale sui Modelli 231 e sui sistemi di compliance aziendale alla luce delle novità introdotte dal DDL sull'AI*, in *Giurisprudenza Penale Web*, 6, 2025. L'Autore, sotto altra angolazione, si è occupato dell'utilizzo dell'intelligenza artificiale quale mezzo per la commissione di ulteriori reati in grado di impegnare la responsabilità dell'ente, nonché quale strumento particolarmente insidioso nella commissione di reati presupposto (già inseriti nel catalogo) della responsabilità degli enti. Nel citato contributo si evidenzia, quindi, l'urgenza di adeguare i modelli di *compliance* aziendale, integrando protocolli di controllo sull'uso dell'I.A., la formazione specifica del personale e la verifica periodica dell'efficacia delle misure adottate.

ISSN 2036-7821

Ma vi è un altro interrogativo speculare e consequenziale al primo.
Partendo dalla premessa, che tali tecnologie hanno la pretesa di ottemperare ai paradigmi dell'esattezza e di maggiore effettività, è indubbio come le stesse siano destinate a incidere progressivamente sul formarsi di *best practices* e, dunque, sulla standardizzazione dei sistemi di *compliance*, quantomeno in alcuni settori.
Come dovranno essere valutati in fase di giudizio di validazione quei protocolli cautelari e più in generale quei modelli organizzativi, che non si siano avvalsi delle nuove misure tecnologiche, ovvero abbiano adottano sistemi informativi, considerati dalla disciplina di settore meno performanti[21] ?.
È indiscutibile che i recenti sistemi informatici possono essere impiegati per perfezionare in particolare, l'attività di mappatura dei rischi che necessita dell'utilizzo di modelli matematici e statistici di matrice predittiva per impedire o ridurre la possibilità di commettere reati
Posto che i modelli matematici e statistici sono parte integrante della struttura ingegneristica dell'intelligenza artificiale, e posto che il modello "matematico" apprestato da un *software,* pur programmato con dati iniziali, è capace di auto-apprendere nuove informazioni e strategie, sarebbe utile il recepimento di quel modello per consentire un continuo adattamento delle aree di rischio (inizialmente mappate), sia ai mutamenti normativi (si pensi all'introduzione di nuovi reati presupposto) sia ai mutamenti che possono involgere i vertici societari, tali da richiedere la revisione dell'organigramma aziendale e del mansionario.
E tuttavia non si può sottacere il "risvolto della medaglia" del soccorso fornito dall'impiego dei *software* (programmati dal creatore) nella selezione delle aree di rischio: ovverosia il pericolo che il sistema predisponga misure analoghe o similari in contesti societari simili *ma non identici*, così trascurando le specificità di ciascuna azienda, che rischierebbero di rimanere prive di adeguati presidi cautelari.
A ciò si aggiunga un ulteriore rilievo: la "esattezza" e la "puntualità" delle scelte operative compiute dai sistemi di I.A. dipende da molte variabili (a cominciare dal tipo di dati elaborati) e non è esente da "distorsioni", perché comunque l'algoritmo è impostato su avvenimenti del passato.
Tra questi pre-giudizi rientra anche il c.d. *automation bias*, cioè la convinzione che la soluzione elaborata dalla macchina sia per ciò solo l'unica corretta, con la necessaria avvertenza, che questa convinzione può essere tanto più fuorviante, se si considera che il risultato atteso riguarda la prevenzione di un comportamento descritto da norme penali soggette a interpretazione e, in molti casi, formulate in modo poco chiaro, se non indeterminato[22].

[21] Su tale specifico punto, si rinvia a N. SELVAGGI, *Dimensione tecnologica e compliance penale*, in L. MARAFIOTI, *Dimensione tecnologica e prova penale*, Giappichelli, 2019, p. 223 ss.
[22] A.NISCO, op.loc.cit

ISSN 2036-7821

Da un lato, dunque, occorre rammentare il problema – più generale – della falsa neutralità dei sistemi predittivi basati sull'IA, che attiene al fondamento epistemologico della predizione algoritmica; dall'altro lato, sia pure in un contesto privato (qual è l'organizzazione aziendale), si ripropone il dubbio inerente alla compatibilità tra un sistema predittivo di tipo informatico e un comportamento contrario a una norma, da giudicare illecito (o comunque deviante) solo al termine di un procedimento logico di sussunzione entro una fattispecie, che la macchina può emulare – o anche aiutare ad impostare –, ma non risolvere compiutamente[23].
È fuori discussione l'utilità dell'I.A. quale guida nel sistema predittivo dei rischi (e della loro relativa mappatura), ed in particolare nei processi di valutazione dei rischi-reato; fermo restando che l'I.A. non può sostituire il giudizio umano, ma al più potenziarlo, supportarlo e renderlo più efficace e nel complesso decisamente più efficiente.
Anche se va notato come la potenziale e tendenziale riduzione degli errori operativi cui va incontro un sistema operativo informatico non è sinonimo di infallibilità del sistema preventivo-predittivo, ben potendo palesarsi disfunzioni in fase di funzionamento dei presidi di controllo, come pure a monte in sede di analisi del rischio.
Vi è infatti che, una volta immessa una determinata tecnologia nell'organizzazione, il suo fallimento concorre al difetto organizzativo, il quale, a sua volta, potrà risalire anche alla fase di progettazione del sistema informatico di gestione del rischio.
In questo modo, però, la comprensione del difetto organizzativo non rientra più nella "signoria" del sinolo ente, che tendenzialmente si è affidato a operatori altamente specializzati.
È indubbio quindi come la "*compliance* digitale" è destinata ad incidere anche sulla logica della auto-normazione, che presiede al concetto di compliance e che caratterizza, in larga parte, anche il sistema 231, mettendo a rischio, con ciò, il fondamento stesso del giudizio di rimproverabilità nei confronti dell'ente.
E tuttavia, per preservare il paradigma di una responsabilità per fatto proprio dell'ente, imputabile per *deficit* d'organizzazione (o per cola da disorganizzazione), si potrebbe anche ipotizzare un obbligo a carico dell'attore collettivo di includere figure professionali esperte, preposte alle doverose verifiche circa l'esatto funzionamento del *software*, prima che si dia corso alla sua installazione.
Ovviamente ai medesimi esperti competerebbe l'obbligo di vigilare e monitorare costantemente il regolare funzionamento dei programmi informatici inizialmente installati dal programmatore.
È innegabile che anche i programmatori o gli informatici più esperti non sempre sono pronti a fornire una spiegazione esaustiva del difetto organizzativo eventualmente

[23] A.NISCO, op.loc.cit.

ISSN 2036-7821

occorso a causa – o nonostante – l'impiego della tecnologia.
Come noto, uno dei principali problemi dell'IA è la trasparenza delle decisioni, la cui imprevedibilità è insita nella stessa creazione di tali sistemi: neanche il programmatore (o comunque il creatore) è in grado di prevedere compiutamente le "decisioni" della macchina artificiale; si è parlato, al riguardo, di una «imprevedibilità prevedibile»[24].
A fronte di tali criticità (non evitabili in via aprioristica) insiti nell'utilizzo dell'I.A. nella compliance aziendale, è bene interrogarsi sulle conseguenze che possano discendere, in caso di commissione di un reato presupposto, dalla dotazione di un modello 231 creato dall'intelligenza artificiale e adottato dalla società.
Per rispondere a tale ultimo quesito, occorre comprendere se l'accertamento della colpa d'organizzazione, comprovata dall'inidoneità del modello, presupponga una valutazione in termini di esigibilità di una condotta diversa.
L'assunto per cui il Giudice deve idealmente porsi al momento in cui il reato è stato commesso e sostituire il modello vigente con quello virtuoso deve, infatti, tenere sempre in debita considerazione lo stato di conoscenze tecniche e scientifiche al momento in cui è stata attuata la condotta.
Tale giudizio controfattuale, non può essere influenzato dalla conoscenza delle conseguenze della condotta posta in essere, rilevate con il senno di poi.
Pertanto, in caso di modello organizzativo redatto tramite un sistema di intelligenza artificiale, bisognerà comprendere se la società che lo ha programmato rivesta una posizione di garanzia sulla corretta gestione del software.
A tal fine, assumerà fondamentale rilievo il contratto di cessione del *software*. Se tale rapporto giuridico esclude qualsivoglia ingerenza da parte del venditore, ricadrà sull'ente che lo ha acquistato il dovere di riprogrammarlo all'occorrenza o, comunque, curarne gli aggiornamenti e vigilare sullo stesso.
Diversamente, se il dominio sulla macchina artificiale rimanesse in capo al venditore, ne deriverebbe il problema di individuare una residuale responsabilità "penale" in capo all'ente per il reato presupposto commesso da un proprio dipendente.
In siffatta evenienza si potrebbe escludere la colpevolezza della società acquirente del *software* per inesigibilità di una condotta alternativa, tenuto conto delle conoscenze tecniche e scientifiche del momento in cui il predetto *software* è stato acquistato sul mercato ed implementato nella *compliance* aziendale, proprio al fine di dotarsi di un'organizzazione efficace in ottica di prevenzione dei reati.
Potrebbe essere semmai rimproverabile all'ente *una nuova ipotesi* di colpa d'organizzazione (*sub specie* di *culpa in vigilando*), per l'omessa predisposizione di un *team* specializzato nell'utilizzo e nella programmazione dei sistemi di intelligenza artificiale.

[24] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762.

ISSN 2036-7821

Tuttavia, proprio in considerazione delle capacità di auto-apprendimento della macchina, anche dagli specialisti del settore non si potrebbe pretendere alcuna condotta impeditiva, rispetto alla verificazione del reato presupposto

Apparirebbe, dunque, come una forzatura l'attribuzione alla società che si è dotata di tale software di una responsabilità da colpa organizzativa per non aver predisposto un modello organizzativo idoneo a prevenire il reato commesso.

De lege ferenda, per superare tali questioni dogmatiche, si potrebbe prevedere una speciale esenzione (o robusta attenuazione) della responsabilità per tutti quegli enti che adottino il modello organizzativo (e costantemente ne curino l'aggiornamento) così come ideato ed elaborato da un *software* frutto di alta tecnologia, programmato appositamente per la mappatura delle aree di rischio in cui la medesima società frequentemente opera.

Si farebbe così largo l'idea di un "agente modello collettivo [algoritmico]"[25], *tecnologicamente avveduto*, senza tuttavia che possa dirsi annullato il rischio di identificazione di uno *standard* di "organizzazione modello" di natura esclusivamente artificiale e suscettibile di implementazioni attraverso il sistema degli algoritmi.

Per vero, voci autorevoli in dottrina[26] hanno immaginato *una responsabilità esclusiva ed autonoma della persona giuridica per il fatto dell'agente artificiale*, ove la fattispecie concreta si concretizzasse in un'organizzazione complessa; e questa sarebbe l'unica possibilità per l'ordinamento di garantire un meccanismo di *accountability* per l'azione di un'intelligenza artificiale.

L'imputazione assumerebbe "una forma sincopata", in cui la condotta della persona fisica funge solo da giunto elastico oggettivo tra il fatto dell'algoritmo e il centro di interesse finale rappresentano dall'ente [27] : una imputazione concettualmente vicariale, ma autonoma ed esclusivamente centrata sull'organizzazione[28].

La diligente organizzazione dell'ente nella gestione di un'AI si presta in effetti a garantire un migliore *enforcement*, tutte le volte in cui l'ente non solo dispone di un modello organizzativo, ma questo si basi su idoneo *know how* operativo e siano dotati di apparati di ricerca e sorveglianza interna che potrebbero tempestivamente attivarsi per prevenire il deragliamento dell'AI o impedire che il danno sia portato a

[25] L'espressione è di A. NISCO, *Riflessi della compliance digitale in ambito 231*, cit., p. 10.

[26] F. CONSULICH, *Flash offenders. Le prospettive di acconuntability penale nel contrasto alle intelligenze artificiale devianti*, in *Riv.it.dir.proc.pen.*, 2022, 3, p. 1042.

[27] Già la dottrina americana ha rilevato come la sostituzione di operatori umani con agenti artificiali possa implicare l'imputazione all'ente di eventi prodotti dal malfunzionamento della macchina, cfr. M.E. DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, in 98 *North Carolina Law Review*, 2020, p 898 ss.

[28] F. CONSULICH, *Il principio di autonomia della responsabilità dell'ente. Prospettive di riforma dell'art. 8*, in *Rivista 231*, 2018, 4, p. 197 ss.

ISSN 2036-7821

conseguenze ulteriori[29].

3. La tesi di Piergallini e la responsabilità diretta della macchina artificiale.

Altra e differente tematica attiene invece alla possibilità, in caso di verificazione del reato presupposto 231, di prevedere che, in luogo della responsabilità dell'ente che ha adottato il modello predisposto dal *software,* la responsabilità *ex delicto* sia "translata" sull'agente artificiale.

Si tratta, in buona sostanza, di riflettere sulla possibilità di coinvolgere la stessa macchina nel circuito punitivo, muovendo dai rilievi sopra operati in punto di classificazione dell'IA come "sistema forte", che, per quel che qui rileva, consente di stabilire un sostanziale *isomorfismo* tra uomo e macchina artificiale, intesa quest'ultima, in una prospettiva c.d funzionalista,[30] alla stregua dell'*homo naturalis*,

In disparte le note pregiudiziali ontologiche (secondo cui il diritto penale si rivolgerebbe solo alle persone fisiche, in forza del principio della "personalità" della responsabilità penale), potrebbe non essere così irrealistico e astratto immaginare una responsabilità amministrativa da reato dell'agente artificiale, "reo" di aver elaborato un protocollo cautelare inidoneo ad escludere la colpa organizzativa dell'ente che quei protocolli ha implementato.

Vengono così in rilievo tre possibili orientamenti[31], già scrutinati con riguardo al tema dell'eventuale responsabilità *penale* del c.d. *homunculus mechanicus* e che – a nostro avviso- potrebbero essere analogicamente estesi, laddove si volesse allargare il *focus* all'eventuale responsabilità amministrativa da reato del soggetto artificiale.

(i) Un primo orientamento punta al riconoscimento *solo* della personalità giuridica della macchina 'pensante', *rectius*: una personalità elettronica, giuridicamente

[29] La prevenzione della criminalità da parte dell'I.A. già passa per i sistemi di "*Criminal Compliance*" basati sui c.d. *Digital Compliance Tools*, i quali consentono il monitoraggio completo delle comunicazioni interne all'azienda con l'obiettivo di "segnalare" le interazioni sospette e quindi di sottoporle a ulteriori indagini. Sul tema, si rinvia ad A. SCHEMMEL - A. DIETZEN, *Effective Corporate Governance by Legal Tech & Digital Compliance*, in S. BREIDENBACH - F . G LATZ (a cura di), *Rechtshandbuch Legal Tech*, München, 2021, p. 167 ss.. A livello europeo, gli artt. 17 e 48 della direttiva MiFID 2 impongono, rispettivamente alle imprese di investimento e alle sedi di negoziazione, di predisporre sistemi di controllo idonei a prevenire gli effetti collaterali indotti sui mercati dalle negoziazioni algoritmiche. Sul punto anche T. CUK - A. VAN WAEYENBERGE, *European Legal Framework for Algorithmic and High Frequency Trading (Mifid 2 and MAR) A Global Approach to Managing the Risks of the Modern Trading Paradigm*, 8 giungo 2020, Available at SSRN: *https://ssrn.com/abstract=3622307* or *http://dx.doi.org/10.2139/ssrn.3622307*, p. 151 ss..

Sull'importanza di un approccio di tipo precauzionale, con adozione di efficienti circuit breakers, supervisioni interne alle imprese di investimento e altri meccanismi in grado di rallentare le transazioni o bloccarne del tutto ove siano indentificati profili di rischio, nonché di un regime di preapprovazione dei sistemi di intelligenza artificiale prima che vengano impiegati nel mercato H. ALLEN, *Driverless Finance,* in 10 *Harv. Bus. L. Rev.*, 2020, 161, p. 196.

[30] L'argomento del c.d. "diritto penale robotico" è stato autorevolmente trattato da C. PIERGALLINI, *Intelligenza artificiale: da "mezzo" ad 'autore" del reato* ?, in *Riv.it.dir.proc. pen*., 2020, 4, p. 1763 ss.

[31] C. PIERGALLINI, *Intelligenza artificiale: da "mezzo" ad 'autore" del reato* ?, in op.loc.ult.cit.

ISSN 2036-7821

rilevante: per questa via, la macchina verrebbe concepita come un autonomo centro di interessi e di imputazione[32], destinato a rispondere dei danni provocati, all'esito ovviamente di un preliminare processo di riconoscimento (di matrice antropomorfica) di tali macchine artificiali sovrapponibile alla stessa certificazione dello *status* delle entità collettive.

Si potrebbe pensare alla creazione di uno specifico registro, nel quale iscrivere determinate macchine "pensanti" con la specificazione delle rispettive caratteristiche tecniche, assegnando loro un numero identificativo (una sorta di carta di identità elettronica).

Questo nuovo ed ipermoderno-soggetto dovrebbe poi essere affiancato da un fondo patrimoniale per rispondere, in proprio, delle obbligazioni: un fondo che dovrebbe essere alimentato, pro-quota, da coloro che, a vario titolo, sono intervenuti nella fase della 'creazione', gestione ed uso del prodotto.

Per finire, si potrebbero prevedere forme di assicurazione obbligatoria. Naturalmente, questa nuova soggettività sarebbe altresì destinataria dei diritti e delle forme di tutela riservate dall'ordinamento agli enti collettivi, da aggiornare in virtù della peculiare natura elettronica del soggetto artificiale. Un soggetto di tal fatta potrebbe risultare destinatario di misure amministrative in senso lato interdittive/incapacitanti e conformative: si pensi, per fare l'esempio più vistoso, ad un provvedimento di 'disattivazione' del sistema, quando lo stesso provochi danni non altrimenti arginabili o reversibili.

(ii) Un secondo orientamento[33], partendo dal riconoscimento di un'autonoma personalità giuridica al soggetto elettronico, perviene ad ipotizzare un timido e pioneristico riconoscimento, anche alla macchina, di una embrionale forma di soggettività penalistica. Vi è tuttavia da precisare che simili intendimenti non hanno trovato favorevole accoglienza; la dottrina più accorsata ha mostrato una robusta contrarietà[34], opponendo il seguente argomento ovverossia: per quanto si possa disquisire sulla capacità di auto-apprendimento di una macchina e della sua similarità con l'uomo, questa sarebbe ontologicamente incapace di colpevolezza: non potrebbe, mai e in alcun modo, percepire un rimprovero; una prerogativa, questa, esclusivamente riferibile a soggetti umani. Tale impostazione non delinea compiutamente un modello di responsabilità penale dell'agente artificiale, ma si limita ad evidenziare come, in futuro, sulla scorta di un eventuale "riconoscimento

---

[32] S. BECK, *Sinn und Unsinn von Statusfragenzu Vor und Nachteilen der Einführung einer elektronischen Person,* in E. HILGENDORF - J.P. G ÜNTHER (a cura di), *Robotik und Gesetzgebung,* Baden, 2013, p. 254 ss.

[33] M. SIMMLER - N. M ARKWALDER, *Roboter in der Verantwortung? - Zur Neuauflage der Debatte um den funktionalen Schuldbegriff,* in *ZSTW*, 2017, p. 20 ss.

[34] F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. cont.*, 29 settembre 2019, p. 30 ss.; A. CAPPELLINI, *Machina delinquere potest?, Brevi appunti su intelligenza artificiale e responsabilità penale,* in *disCrimen*, 27 marzo 2019, p. 14 ss.

ISSN 2036-7821

sociale", potrebbero non sussistere insuperabili ostacoli al confezionamento di una nuova, tecnologica "soggettività giuridica", verso la quale indirizzare un autonomo sistema punitivo.

Vero è, che per le persone artificiali o, meglio, per le c.d. responsabilità artificiali *l'identificazione della colpa* è tuttavia operazione dogmatica più complessa e per certi versi invertita, nel senso che, non è la persona giuridica che non ha impedito la colpa della persona fisica, ma è la persona fisica che non ha impedito la colpa della persona artificiale.

La colpa, tanto della persona fisica quanto della persona artificiale, consisterebbe nella incapacità di impedire la realizzazione di illeciti, che invece, entrambi, avrebbero potuto impedire; la persona giuridica attraverso regole cautelari che "plasmano" la colpa di organizzazione; la persona fisica attraverso una congerie di *standard* costruttivi, protocolli, cautele che avrebbero potuto impedire la generazione artificiale di condotte illecite: una colpa di programmazione, assai simile alla colpa di organizzazione.

L'intelligenza umana deve competere con l'intelligenza artificiale che, a tutto voler concedere, è più veloce di quella umana, come se nella partita del *facere illecito* e della guida lecita si misurassero due giocatori con livelli di intelligenza differenti[35]. La fraudolenta elusione – sintagma usato nel paradigma del modello ascritto della responsabilità 231 –avrebbe margini di probabilità altissimi, tanto quanto la manifesta superiore capacità dell'I.A. di aggirare le regole seppellite nel *black box* dell'algoritmo inziale: i metodi di contenimento, i metodi di incentivazione, l'impedimento dello sviluppo, il controllo sul surplus di modernità nelle forme dell'"obsolescenza programmata", i t*ripwires*, ossia meccanismi di diagnostica random, ad insaputa del sistema potrebbero non essere sufficienti e l'IA potrebbe sapientemente ingannare i "guardiani umani"[36].

---

[35] V.B. MUSCATIELLO, *Ex Machina. Il diritto penale, all'improvviso,* in *L'Ircocervo*, 2024, 1, pp. 20 ss..

[36] V.B. MUSCATIELLO, *Ex Machina. Il diritto penale, all'improvviso*, op loc. cit.; in particolare l'Autore ritiene che la pre-colpa, l'idea cioè di una governance per prevenire attività illecite, sconta un'evidente vulnerabilità logico dommatica: la oggettiva difficoltà, forse persino impossibilità, per l'intelligenza umana di esercitare la sorveglianza sui sistemi di IA, in particolare su quelli ad apprendimento automatico, capaci di compiere operazioni e prendere decisioni a tal punto complesse da sorpassare la capacità umana di controllo, "*eccedendo non solo i loro limiti cognitivi, ma anche quelli temporali nell'accedere alle informazioni, elaborarle e prendere decisioni*". Il giudizio umano è, in sostanza, largamente sottodimensionato rispetto alla complessità dei modelli che guidano le scelte dei sistemi artificiali, il che rende poco plausibile pensare che una intelligenza umana, per quanto vivace, possa evitare l'elusione fraudolenta di un modello posta in essere da una intelligenza artificiale; che possa prevedere e scoprire l'astuzia fraudolenta di un sistema straordinariamente più agile e veloce; che un controllore umano possa controllare un controllato artificiale. Anche quando l'uomo dovesse mantenere formalmente il controllo sulla decisione finale, la possibilità che egli sia in grado di entrare efficacemente nel merito della decisione si tradurrebbe in un'ipotesi remota, e più verosimilmente, il supervisore umano tenderà a fare affidamento sulle decisioni di un sistema di IA, tanto più se questo è stato certificato, a meno che non abbia motivi specifici per ritenere che esso sia malfunzionante, o che non sia in grado di incorporare nella sua valutazione ulteriori elementi, esterni al sistema stesso.

ISSN 2036-7821

(iii) Un terzo orientamento[37] poggia infine sul *pieno riconoscimento* di *una responsabilità penale ascrivibile alla nuova persona giuridica* (la c.d. *machina artificialis*), invocando i seguenti argomenti "paralleli" con il paradigma della responsabilità penale delle persone giuridiche, ovverosia che:
- anche le macchine sono certamente capaci di azioni penalmente rilevanti (si pensi al movimento di un braccio robotico);
- i sistemi di IA vanterebbero tanto l'elemento della cognizione che quello della volizione. Se è vero che la cognizione umana è fondata sulla capacità di acquisire dati dall'ambiente esterno e di rielaborarli, previa una loro rappresentazione, una simile abilità sarebbe senz'altro da riconoscere all'IA, che, è pacifico, auto-apprende sulla scorta degli stimoli che provengono dall'esterno. Quanto alla volizione poi si può rintracciarla anche nell'IA, atteso che si danno macchine idonee ad apprezzare le probabilità di verificazione di un evento e di agire di conseguenza. Circa la *negligence*, sarebbe possibile reperirla nelle macchine, forgiando una inedita versione, di stampo metaforico, di "agente-modello", alla cui stregua determinare se la macchina che ha agito si sia o meno discostata da ciò che avrebbe fatto una "IA-modello" munita delle pregresse esperienze, degli stessi dati esaminati e della identica capacità di auto-apprendimento della macchina-concreta;
- che i *software* potrebbero tranquillamente essere destinatari di pene riabilitative o privative della libertà di agire.

Questo terzo orientamento è stato avversato dalla dottrina più avveduta[38], secondo cui è "malfermo" il parallelismo tra la responsabilità penale diretta dell'agente artificiale e la responsabilità penale della *societas*, posto che quest'ultima, sia che la si interpreti in guisa di uno "schermo", sia alla stregua di un "organo", esiste nella realtà giuridica e sociale, ma è animata, naturalisticamente e spiritualmente, dagli uomini che le danno vita: i precetti che le vengono rivolti sono diretti ai soggetti che l'hanno creata e che la rappresentano, condizionandone i comportamenti (specie in vista della dotazione di un'adeguata organizzazione preventiva dei reati). I sistemi di IA possiedono, per contro, una 'fisicità', in quanto vengono concepiti dal loro ideatore, distaccandosi da esso, e agiscono autonomamente.

Le macchine difettano della *suitas* visto esse non agiscono ma "sono agite". A tal riguardo, si è acutamente rilevato in dottrina[39]: "*anche se è vero che rielabora i dati della realtà esterna (sì da conoscerla) a scopo di auto- apprendimento, le decisioni che partorisce sono una conseguenza immediata e diretta riferibile all'algoritmo che le governa, che, però, è sprovvisto della libertà di autodeterminazione che contraddistingue l'esperienza degli esseri umani. Il decision making non è frutto di una scelta ma è l'obbligata risultante dei*

[37] G. HALLEVY, *The Criminal Liability of Articial Intelligence Entities. From Science Fiction to Legal Social Control*, in *Akron. Intell. Prop.J.*, 2010, p. 171 ss.; ID, *Liability for Crimes Involving Artificial Intelligence Systems, Springer, 2015.*
[38] C. PIERGALLINI, op.loc.cit., p. 1768 ss.
[39] C. PIERGALLINI, op.loc.cit., p. 1769.

ISSN 2036-7821

meccanismi oscuri che alimentano l'algoritmo, a sua volta forgiato dall'uomo. Il vero motore dell'azione va pur sempre ricercato a monte, in colui che elabora l'algoritmo: quel che accade "dopo" si espone ad una imprevedibilità necessitata, non altrimenti governabile, che non può in alcun modo essere incasellata come manifestazione di un agire intelligentemente gestito, perché è carente del requisito costitutivo della libertà di autodeterminazione".

Peraltro, le macchine difettano della possibilità di essere rimproverabili, perché sprovviste della capacità di scelta, del libero arbitrio; e nessuna pena potrebbe essere efficacemente "riabilitativa" per l'agente artificiale in quanto la rieducazione si fonda pur sempre sulla possibilità che la sanzione irrogata aiuti a comprendere l'errore commesso per evitare di reiterare nello stesso. Quindi il presupposto è una capacità di autoriflessione che la macchina, ontologicamente, non possiede.

L'applicazione della pena presuppone la attribuibilità psicologica e la rimproverabilità del fatto di reato al soggetto che lo ha posto in essere e la sanzione penale inflittagli deve tendere alla sua rieducazione(art. 27, co. 3, Cost.); ma questo, come si è detto, non può valere per l'agente artificiale.

In definitiva, per la summenzionata dottrina (Piergallini), si deve riconoscere che "*allo stato, machina artificialis delinquere et puniti non potest*".

4. Soluzioni de lege ferenda per promuovere l'impiego dell'I.A. nella costruzione dei M.O.G.

Occorre prendere atto come molteplici potrebbero essere gli apporti futuri dell'I.A. ed in particolare quelli delle c.d. "macchine pensanti" che nell'avvenire potrebbero diventare operative.

Tali "macchine umane" per le loro immense potenzialità di immagazzinamento dei dati provenienti dall'esterno potrebbero elaborare complessi algoritmi utili alla predisposizione di modelli pilota differenziati in ragione delle specifiche attività economiche. Detti modelli verrebbero così integralmente concepiti dall'I.A e dovrebbero a loro volta -prima di diventare pubblicamente fruibili- essere previamente sottoposti alla validazione pubblica (ad esempio ministeriale).

Dopo la predetta validazione pubblica, si potrebbe prevedere una presunzione di idoneità *iuris tantum* dei modelli organizzativi che l'ente adotta in piena conformità con quelli "pilota" pubblicamente validati; cosicchè, in sede giudiziale tale presunzione relativa di idoneità potrebbe essere vinta solo tramite la c.d. motivazione rafforzata.

Entrando più nel dettaglio, si potrebbe pensare di utilizzare le immense potenzialità dei *software* (o dell'agente artificiale) per elaborare "modelli pilota" o "protocolli pilota", raccogliendo e catalogando la massa di dati provenienti dalle norme tecniche

ISSN 2036-7821

(norme ISO e assimilabili) o le linee guida più accreditate nei diversi settori[40].
In buona sostanza, si potrebbe affidare alle c.d. macchine pensanti il delicato compito di una maggiore positivizzazione delle cautele, attraverso il ricorso ad un procedimento di elaborazione in grado di coinvolgere più professionalità specifiche i cui prodotti dovrebbero recepire le c.d. *best practice.*
I passaggi operativi potrebbero essere i seguenti: i) le associazioni di categoria dovrebbero contribuire all'emersione, per omogenei settori di attività, delle cautele adottate dagli enti, favorendo la circolazione, in modo "protetto", dei contenuti della Parte speciale dei modelli organizzativi; ii) si dovrebbe poi procedere a una revisione critica delle misure di *compliance* esistente, operata con il supporto di 'saperi' esperti (aziendalistici, societari e giuridico-penalistici) e tenendo conto sia delle evidenze scientifiche disponibili, sia degli «indici qualitativi» individuati dalla giurisprudenza; iii) una volta raccolti tutti i dati, alla macchina pensante sarebbe affidato il compito di concepire lo specifico modello in ragione del singolo settore economico, nonché di occuparsi un'attività di costante aggiornamento che consentirebbe non solo di individuare le *best practice*, ma anche i più comuni errori.
Come si è detto, occorrerebbe altresì *de lege ferenda* – prevedere una presunzione *iuris tantum* di idoneità, per cui, se il modello in concreto implementato nella società incolpata è conforme a quello "pilota" (concepito dall'I.A.), il giudice che rilevi, ciononostante, profili di inidoneità dovrà darne conto con motivazione rafforzata.
In sede di attività di raccolta della massa dati, il *software* potrebbe anche immagazzinare ed elaborare tramite algoritmi le informazioni provenienti anche da soggetti "terzi" con esperienza tecnica sul campo (avvocati, magistrati, accademici), costituiti in tavoli/gruppi di lavoro per settori di attività e tipologie di imprese, che siano così in grado di offrire un contributo qualificato in sede di ricognizione, definizione e revisione delle cautele pilota.
In tale contesto, l'I.A. potrebbe anche valutare dati e informazioni provenienti da alcune authority, sia rispetto alla elaborazione di protocolli specifici per aree/reati presupposto (ad esempio l'ANAC per la prevenzione delle fattispecie di corruzione), sia, sulla base della relativa "competenza"/area di attività.
Una possibile soluzione potrebbe essere la previsione di un *iter* per la validazione preventiva del singolo modello organizzativo concepito dal singolo *software* di cui la

[40] Un esempio, nel nostro ordinamento, è rappresentato dall'art. 30 del d.lgs. n. 81/2008, che specifica le componenti essenziali del modello organizzativo per la prevenzione dei reati colposi che ledono la salute e la sicurezza dei lavoratori. Il comma 5 dell'art. 30 citato prevede, in particolare, una presunzione di adeguatezza del modello conforme ad alcune regole tecniche/standard nazionali o internazionali (le Linee guida UNI-INAIL per un sistema di gestione della salute e sicurezza sul lavoro del 28 settembre 2001 o il British Standard OHSAS 18001:2007; oppure le ISO 45001:2018 – Sistemi di gestione per la salute e sicurezza sul lavoro. Presunzioni di conformità di questo tenore potrebbero essere estese ai modelli pilota adottati in altri settori operativi esposti a rischio penale: ambiente (ISO 14001), corruzione (ISO 37001), criminalità informatica (ISO 27001).

ISSN 2036-7821

specifica azienda si sia dotato; in questa seconda evenienza non vi sarebbe una fase di accreditamento dell'unico modello pilota, ma si potrebbe pensare ad una validazione del singolo modello realizzato dalla macchina pensante (o dall'agente artificiale) al servizio di quella specifica società e prima della sua implementazione nel contesto aziendale.

In siffatta evenienza diventerebbe esiziale la disciplina puntuale della fase di accreditamento/validazione, "prima e fuori" del processo, del modello organizzativo concepito dall'I.A.; si tratterebbe quindi di prevedere una sorta di certificazione o attestazione preventiva del sistema di *compliance* penale ad opera di organismi abilitati, vale a dire di soggetti (privati e pubblici) investiti del potere di validare la regolare conformità dei contenuti del modello scrutinato.

Si deve tuttavia tener conto come l'analoga prospettiva di una fase di accreditamento preventivo del modello (concepito dall'uomo e non dalla macchina pensante) non ha suscitato particolare entusiasmo tra taluni studiosi della tematica[41] che hanno avanzato alcune riserve: in particolare è stata messa in dubbio la praticabilità di un'attestazione concernente la conformità di misure organizzativo-gestionali in quanto condurrebbe alla "degiurisdizionalizzazione" del sindacato sulla idoneità del modello, in un settore normativo a forte impronta pubblicistica.

E' stata poi rimarcata l'intangibilità del libero convincimento del giudice, non conculcabile in un settore normativo a forte impronta pubblicistica; è stata altresì rimarcata l'intangibilità del libero convincimento del giudice, non conculcabile attraverso prove legali vincolanti; è stata poi fatta notare la scarsa compatibilità del carattere "istantaneo" del giudizio di attestazione con la necessaria dimensione "diacronica" e dinamica del modello e, più in generale, della realtà aziendale (per sua natura in continuo divenire).

Infine è stata sottolineata l'impossibilità di scindere la valutazione sull'adeguatezza astratta del modello da quella concernente la sua concreta effettività e l'assenza di un chiaro standard tecnico di riferimento, analiticamente definito, sui cui misurare il vaglio di conformità[42].

In particolare, la riserva più significativa si è incentrata sulla circostanza che il giudizio di un attestatore non potrebbe che essere formulato sul modello preventivo

[41] G.M. F LICK, *Le prospettive di modifica del d.lg. n. 231/2001, in materia di responsabilità amministrativa degli enti: un rimedio peggiore del male?,* in *Cass. pen.*, 2010, p. 4032; nonché G.DE VERO, *Il progetto di modifica della responsabilità degli enti tra originarie e nuove aporie*, in *Dir. pen. proc.*, 2011, n. 10, p. 1140; F. MUCCIARELLI, *Una progettata modifica al D.Lgs. n. 231/01: la certificazione del modello come causa di esclusione della responsabilità*, in *Le società*, 2010, n. 10, p. 1247 ss.; S. MANACORDA, *L'idoneità preventiva dei modelli di organizzazione nella responsabilità da reato degli enti: analisi critica e linee evolutive*, in *Riv. trim. dir. pen. ec.*, 2017, 1-2, p. 100 ss.; R. SABIA, *Responsabilità da reato degli enti e paradigmi di validazione dei modelli organizzativi,* op. cit., p. 157 ss..

[42] Qualcosa di simile alla certificazione preventiva è stato positivizzato nel settore nevralgico della sicurezza sul lavoro. Ci riferiamo alla denominata "asseverazione" dell'adozione ed efficace attuazione dei modelli di organizzazione e gestione della sicurezza di cui all'art. 30 del d.lgs. n. 81/2008, asseverazione prevista dall'art. 51, comma 3-*bis*, d.lgs. n. 81/2008.

ISSN 2036-7821

"allo stato degli atti", anziché su quello attuato nel corso del tempo.
Le medesime riserve potrebbero essere avanzate rispetto ad un *compliance programs* elaborato da un software in uso ad una specifica azienda e sottoposto per la validazione preventiva ad un organo pubblico/privato che sia abilitato e sia terzo rispetto al mondo imprenditoriale.
In effetti lo scrutinio *post delictum* dell'organo giudicante si muove inevitabilmente su un terreno diverso, posto che si premura di confrontare il fatto storico contestualizzato con il modello comportamentale allora impiantato dall'ente. E la verifica di corrispondenza tra il modello attestato e quello reso operativo in concreto non può che restare all'interno della cornice decisionale dell'Autorità giudiziaria, sulla base delle risultanze acquisite in sede di indagini preliminari ed in sede di istruttoria dibattimentale.
Più convincente ci sembra l'approccio *de lege ferenda* di quella parte autorevole della dottrina[43], che suggerisce – nel caso di implementazione di modello validato da un ente terzo e abilitato- l'introduzione di misure premiali in favore dell'ente: quali l'esenzione dalle più impattanti misure interdittive per la società sia in fase cautelare sia all'esito del giudizio di merito, nonché la possibilità per l'ente di invocare la messa alla prova o la non punibilità per tenuità del fatto (alla stregua di quanto previsto dall'art. 131-*bis* in caso di illeciti penali).

5. L'inedita dimensione antropica dell'imputazione dei "reati presupposto 231": la responsabilità penale dei programmatori/produttori dei software utilizzati nella redazione dei protocolli cautelari.

Una volta ipoteticamente esclusa la responsabilità dell'ente "avveduto" (che si è dotato di un software in astratto evoluto ed efficiente, che però -alla prova dei fatti- si è rivelato inidoneo a neutralizzare il rischio di consumazione degli illeciti presupposto 231), occorre indagare l'ulteriore problematica dell'eventuale responsabilità penale "indiretta" (aggiuntiva rispetto a quella diretta degli autori materiali) a carico degli *sviluppatori di software* con riguardo ai reati presupposto che sono stati consumati in seno all'ente cessionario del *software*.
In altri termini, ove il modello organizzativo 231 fosse il frutto di algoritmi elaborati dall'I.A., e ciò nondimeno il reato presupposto si fosse verificato, da un lato si potrebbe ipotizzare la non punibilità della persona giuridica, e dall'altro si potrebbe ricorrere ad un meccanismo di responsabilità vicariale o diretta dell'essere umano (il programmatore o produttore) per aver concepito algoritmi o un *software* che ha prodotto un modello organizzativo inidoneo a perseguire *la corretta gestione del rischio* (ossia l'adozione di protocolli e procedure atti a mappare i rischi derivanti dall'attività di impresa, e a monitorare l'emersione di nuovi rischi).

[43] V. MONGILLO, op.loc.cit.

ISSN 2036-7821

In altri termini, il rimprovero di carattere squisitamente colposo da muovere al programmatore sarebbe quello di non aver predisposto i presidi cautelari idonei a neutralizzare o a minimizzare i rischi di verificazione dei reati presupposto 231.
E' indubbio che la responsabilità del produttore non va intesa in senso puramente oggettivo, quale effetto automatico legato all'esistenza del reato presupposto, ma va modulata in base alle categorie penalistiche del dolo o della colpa ed in base ai criteri generali di ascrizione della responsabilità penale.
Il programmatore assume le vesti dell'*employer* e il sistema intelligente sarebbe il suo *employee*, definendo un paradigma basato sulla c.d. conseguenza probabile (*natural-probable-consequence*)[44].
I paradigmi imputativi di responsabilità del programmatore (e del sorvegliante, in ragione della "posizione" dell'agente rispetto alla condotta dell'I.A.) potrebbe diversamente atteggiarsi sia a titolo di dolo che di colpa.
Nella prima ipotesi, quella in cui il *software* è stato volutamente programmato quale strumento (o mezzo) per agevolare -nel corso dell'attività imprenditoriale- la consumazione di reati[45] (si pensi a chi dolosamente diffonde codici *malware* o *algoritmi lesivi*[46] oppure fornisce *istruzioni lesive all'algoritmo* al fine di generare un modello organizzativo 231 inidoneo alla gestione del rischio), non si sarebbero particolari difficoltà dommatiche a ritenere che il programmatore potrebbe essere chiamato a rispondere, a titolo di *concorso doloso*[47], del reato presupposto doloso o

[44] Con riferimento ad un tema analogo, ossia la responsabilità del programmatore per il fatto dannoso dell'I.A., si rinvia a F. CONSULICH, *Flash offenders. Le prospettive di accoununtability penale nel contrasto alle intelligenze artificiale devianti*, in *Riv.it.dir.proc.pen.*, 2022, 3, p. 1031 ss.; in particolare l'Autore evoca lo scenario anglosassone, laddove la responsabilità del predetto programmatore viene inquadrata nel paradigma tradizionale del concorso di persone, per ascrivere al partecipe un fatto non voluto (dunque riconducibile ad una mera *negligence* o al più *recklessness*) realizzatosi nel contesto di un accordo per commettere un diverso reato meno grave (è in sostanza una disciplina strettamente imparentata, dal punto di vista concettuale, al nostro art. 116 c.p.). Proprio il concorso di persone è il contesto in cui più di frequente si ricerca un'adeguata qualificazione del fatto compiuto dall'AI. In particolare, quest'ultima fungerebbe da autore materiale di un reato progettato dalle prime. È sottesa alla riflessione della dottrina che ha studiato la rilevanza giuridica di simili nuovi soggetti la dicotomia tra *homo tecnologicus* (o *homo artificialis*) e *homo naturalis*.

[45] D. COSTA, in *L'impatto dell'Intelligenza Artificiale sui Modelli 231 e sui sistemi di compliance aziendale alla luce delle novità introdotte dal DDL sull'AI*, op. cit., laddove l'I.A. è analizzata quale strumento particolarmente insidioso nella commissione di reati presupposto della responsabilità degli enti, tant'è che si si evidenzia l'urgenza di adeguare i modelli di compliance aziendale, integrando protocolli di controllo sull'uso dell'IA, la formazione specifica del personale e la verifica periodica dell'efficacia delle misure adottate.

[46] In dottrina, si è parlato di responsabilità dolosa del produttore per gli *output* lesivi derivanti da un sistema di I.A. generativa; si rinvia in particolare a B. FRAGASSO, *L'intelligenza artificiale e responsabilità penale,* Giappichelli, 2025, pp. 583 ss.

[47] B. FRAGASSO, op. loc. cit. pp. 587 ss., laddove il tema della responsabilità concorsuale del produttore di sistema di I.A. è stato analizzato più in generale con riguardo al contributo causale offerto alla realizzazione di un qualunque reato da parte dell'utilizzatore (si pensi alla responsabilità concorsuale dei fornitori dei servizi digitali per i contenuti illeciti pubblicati dagli utenti).

ISSN 2036-7821

colposo posto in essere dagli apicali o dai sub-apicali dell'ente complesso[48].
Ne consegue che il produttore/programmatore di un'attività algoritmica inidonea a prevenire il rischio di accadimento dei reati presupposto 231 potrà rispondere a titolo di *concorso doloso* sempre che, oltre a rappresentarsi genericamente il comportamento algoritmico, abbia avuto *contezza specifica del risultato lesivo* del sistema di I.A. (o della forma specifica di offesa).
Si pensi ad esempio a quel programmatore che abbia creato volutamente un *chatbot* generativo di informazioni false poi trasfuse nel modello 231 così da renderlo del tutto inefficace a prevenire le situazioni di rischio-reato e segnatamente la verificazione dei *predicate crimes*. Tale incriminazione a titolo di dolo incontrerà comunque il limite invalicabile rappresentato dall'imprevedibilità e autonomia (ove accertate) dei meccanismi algoritmici,
Più problematico sarebbe l'inquadramento teoretico, nel differente "tipo" del concorso *colposo*, delle condotte degli sviluppatori per la negligente gestione del rischio da intelligenza artificiale e comunque per aver confezionato per imprudenza, imperizia e disavvedutezza un prodotto artificiale "difettoso"[49].
Trovandoci, infatti, al cospetto di un contesto in cui l'evento lesivo costituisce l'esito finale di una lunga serie di "decisioni" algoritmiche, imponderabili ed imprevedibili assunte dal sistema di I.A. (si pensi alla c.d. imprevedibilità algoritmica), il *nucleo del rimprovero* che potrà essere mosso allo sviluppatore non consiste tanto nella realizzazione dell'evento di danno, quanto nell'aver commercializzato un prodotto *intelligente* — di cui ben si conoscono, quantomeno in via generale, le potenzialità lesive — in assenza delle dovute cautele ed accortezze. Pur avvertendo l'esigenza di evitare forme di deresponsabilizzazione, sarà necessaria una limitazione della responsabilità penale, al fine di evitare che il produttore/programmatore sia chiamato a rispondere anche di quegli eventi lesivi derivanti da comportamenti incontrollabili e imprevedibili dell'I.A. (ai fini della nostra ricerca gli eventi lesivi sono quelli coincidenti con la verificazione dei reati presupposto 231 né impediti, né ostacolati dai protocolli cautelari creati dal *software*).
Non è ammissibile ovviamente una forma di responsabilità *in re ipsa*, ossia non basta il semplice fatto di aver programmato e sviluppato un sistema di I.A. per determinare, di per sé, in capo allo sviluppatore, una responsabilità per le decisioni

[48] Si segnala, a tal riguardo, come l'articolo 26, comma 1, del d.d.l. sull'I.A., approvato il 25 giugno 2025 in seconda lettura con modifiche alla Camera (Atto Camera n. 2316) preveda espressamente: i) l'introduzione di una circostanza aggravante comune, qualora il reato sia commesso mediante sistemi di intelligenza artificiale; ii) l'inserimento nel codice penale di una circostanza aggravante a effetto speciale legata all'impiego di sistemi di intelligenza artificiale nella commissione del delitto di attentati contro i diritti politici del cittadino di cui all'art. 294 c.p.; iii) l'introduzione del nuovo reato di illecita diffusione di contenuti generati o manipolati con sistemi di intelligenza artificiale.
[49] F. BASILE – B. FRAGASSO, *Intelligenza artificiale e diritto penale: prove tecniche di convivenza*, in *Discrimen*, 2023, 3-4, p. 762 ss.

ISSN 2036-7821

assunte dall'algoritmo. Non è configurabile quindi una responsabilità penale "oggettiva" e "per risalita" della persona fisica in relazione al "comportamento" di agenti artificiali che non hanno alcuna comprensione simbolica e semantica della realtà umana[50].

È in quest'ottica che, *de jure condendo*, diversi autori[51] hanno proposto l'introduzione di forme anticipate di tutela penale, sotto forma di paradigma imputativi di reati di pericolo astratto che, quantomeno in determinati settori di sviluppo dell'intelligenza artificiale, sanzionino la creazione o il mantenimento di rischi illeciti a beni giuridici di primaria importanza, quali la vita e l'integrità fisica[52].

Non si può tuttavia sottacere che l'introduzione di reati di pericolo astratto — oltre a porsi in uno stato di inevitabile tensione con il principio di offensività — si sia rivelata, in passato, tutt'altro che risolutiva: si pensi, in particolare, alle

[50] La letteratura che esamina i rapporti tra intelligenza artificiale e responsabilità (dal punto di vista giuridico, così come dal punto di vista filosofico) è ormai davvero sterminata. Per l'indicazione dei principali contributi in ambito penalistico, si rinvia a F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione,* in G. BALBI, (a cura di), *Diritto penale e intelligenza artificiale, "Nuovi Scenari"*, Giappichelli, 2023, p. 1 ss.; 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.*, 2024, fasc. 1, p. 287 ss.. Per una riflessione filosofica sui temi della responsabilità è imprescindibile A. MATTHIAS, *The responsibility gap: Ascribing responsibility for the actions of learning automata,* in *Ethics and Information Technology*, 2004, vol. 6, p. 175 ss.

[51] N. AMORE, *L'effetto della robotica e dell'IA nell'imputazione giuridica degli eventi infausti*, in N. AMORE - E. ROSSERO, *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità,* Il Mulino, 2024, p. 249 ss.; F. CONSULICH, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti,* in *Riv. it. dir. proc. pen.*, 2022, 3, p. 1053

[52] In aggiunta alla tematica della tutela penale a presidio dei danni da intelligenza artificiale difettosa o deviata, resta la questione della riconducibilità della responsabilità derivante dall'utilizzo di agenti artificiali alla categoria della responsabilità per danno da prodotto v., nella dottrina civilistica, BERTOLINI A., *Robots as Products: The Case for A Realistic Analysis of Robotic Applications and Liability Rules,* in *Law, Innovation an Technology*, 5(2), 2013, p. 214 ss., p. 235 ss.; più di recente U. R UFFOLO, *Intelligenza artificiale, machine learning e responsabilità da algoritmo,* in *Giur. it.*, 2019, p. 1689 ss.; G. WAGNER, *Produkthaftung für autonome Systemen*, in *AcP*, Bd. 217, 2017, p. 707 ss.; K.A. CHAGAL - FEFERKORN, *Am I An Algorithm or A Product? When Products Liability Should Apply To Alhorithms Decision-Makers,* in *Stanford Law & Policy Rev.*, 30, 2019, p. 61 ss.. Sul versante più prettamente penalistico, si rinvia A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Discrimen*, 2019, p. 1 ss., p. 9 ss.; C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv.it.dir.proc.pen.*, 2020, 4, p. 1745 ss. Più in generale, sui complessi problemi giuridico-penali che solleva la responsabilità per danno da prodotto, vedi, oltre al contributo monografico di C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Milano, 2004; ID, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, in *Dir. pen. proc.*, 2007, p. 1125 ss.; ID, *Colpa e attività produttive: un laboratorio di diritto "cedevole"*, in *Criminalia*, 2014, p. 387 ss.; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori,* in *Riv. trim. dir. pen. econ.*, 2003, p. 1 ss.; D. CASTRONUOVO, *Responsabilità da prodotto e struttura del fatto colposo*, in *Riv.it.dir.proc. pen.*, 2005, p. 301 ss.; nella dottrina tedesca v. L. KHULEN, *Fragen einer strafrechtlichen Produkhaftung*, Heidelberg, 1989; W. HASSEMER, *Produktverantwortung im modernen Strafrecht*, Heidelberg, 1994; K. TIEDEMANN, *Körperverletzung und strafrechtliche Produktverantwortung*, in *FS H.J. Hirsch*, Berlin, 1999, p. 765 ss.; S. GLESS, *Strafrechtliche Produkthaftung*, in *Recht*, 2013, 2, p. 54 ss.; in quella spagnola, si rinvia a S. MIR PUIG, D.M. LUZÓN PEÑA (dirs.), *Responsabilidad de las empresas y sus organós y responsabilidad por el producto*, Barcelona, 1996; M.E. IÑIGO CORROZA, *La responsabilidad penal del fabricante por defectos de sus productos*, Barcelona, 2001; J. BOIX REIG, *Responsabilidad penal por defectos en productos destinados a los consumidores*, Madrid, 2005.

ISSN 2036-7821

contravvenzioni previste dalla disciplina generale sulla sicurezza dei prodotti (vedi art. 112, codice del consumo), che, ad oggi, sono rimaste soltanto sulla carta, complici anche cornici edittali non del tutto adeguate allo scopo di deterrenza.
Il rischio da evitare, allora, è che i nuovi reati si vadano aggiungere a quella pletora di contravvenzioni che abbondano nella legislazione complementare, e che, sulla spinta di esigenze di tutela settoriali, hanno determinato una disordinata evoluzione ipertrofica del diritto penale.
Sicchè sarà indispensabile operare una selezione acuta, rigorosa e improntata ai criteri di precisione ed *extrema ratio*[53] delle condotte degli sviluppatori di *software* da sanzionare penalmente.
Sullo sfondo resta l'altra problematica legata alle ipotesi di concorso a titolo soggettivo eterogeneo e segnatamente riconducibili al paradigma imputativo colposo (per imperizia, imprudenza e negligenza) dei programmatori, rispetto al catalogo dei reati presupposto di matrice tanto dolosa, quanto colposa.
Nella prima evenienza si potrebbe immaginare un concorso colposo del programmatore nel delitto doloso dei soggetti qualificati[54], mentre nella seconda evenienza si potrebbe configurare un concorso colposo nel delitto presupposto colposo riconducibile al cono applicativo della cooperazione colposa ex art. 113 c.p..[55]
In questa sede, si ritiene utile soffermarsi unicamente sulla violazione, *a titolo di colpa,* delle regole cautelari ad opera del programmatore. È fuor di dubbio che le condotte "incaute" e disaccorte di quest'ultimo siano –in astratto- tali da contribuire, quanto meno materialmente, alla realizzazione e messa in commercio del *software,* così *agevolando* o *facilitando* la commissione del reato presupposto, innescando quella progressione criminosa che esita nella condotta illecita altrui (*rectius*: nella verificazione del reato presupposto 231).
Inoltre, sul terreno dell'elemento soggettivo, è altresì ipotizzabile che il programmatore abbia agito nella consapevolezza che quel *software* sarebbe stato successivamente impiegato nella *compliance* aziendale e segnatamente nella fase di

[53] F. BASILE – B. FRAGASSO, *Intelligenza artificiale e diritto penale: prove tecniche di convivenza*, in *Diritto e difesa,* 2023, 3-4, p. 764.

[54] Il nodo della configurabilità del concorso colposo nel reato doloso costituisce senz'altro uno dei punti maggiormente problematici anzitutto sul fronte del fondamento normativo della costruzione teorica; secondo un consistente orientamento dottrinale, il concorso colposo nel fatto doloso non avrebbe alcuna base giuridica, posto che l'art. 110 c.p. e l'art. 113 c.p. legittimerebbero solo le rispettive ipotesi di concorso omogeneo, giungendo anche a consentire il concorso doloso nel fatto colposo. A questo primo argomento, poi, si associa l'ulteriore rilievo per cui l'ordinamento conoscerebbe già le ipotesi di agevolazione colposa, previste in via tassativa nella parte speciale: il che, evidentemente, implicherebbe che una costruzione "di parte generale" non avrebbe altro scopo se non destituirle di ogni fondamento, così G. FIANDACA - E. MUSCO, *Diritto penale. Parte generale*, IX ed., Bologna, 2024, p. 535.

[55] In tema di cooperazione colposa nei complessi organizzati, si rinvia a G. MINICUCCI, *La plurisoggettività nell'agire colposo Una rilettura in chiave normativa. Una rilettura in chiave normativa,* in *Jura,* Edizioni ETS, 2025, pp. 66 ss.

ISSN 2036-7821

valutazione della mappatura di rischi la c.d. attività di *risk assessment*, ineludibile nella costruzione dei protocolli cautelari, secondo la sequenza interattiva tra condotte, in cui il compartecipe è consapevole che la propria condotta accede a quella di altri.

Qui però sorge un altro profilo di problematicità, legato all'eventuale responsabilità penale degli *autori materiali* dei reati presupposto di matrice colposa, che potrebbero invocare -per neutralizzare ogni profilo di colpa a loro ascrivibile- il c.d. principio di (aver fatto) affidamento sull'assenza di vizi del *software* (come progettato e realizzato dal programmatore), con l'unico limite naturale rappresentato da quei casi in cui i difetti di costruzione del programma fossero agevolmente conoscibili da parte dei medesimi agenti.

Potrebbero dunque richiamarsi in questa sede i principi generali in tema di cooperazione colposa laddove la riconoscibilità del difetto del programma elettronico potrebbe fondare un giudizio di rimprovero a carico dell'agente per la mancata adozione di quel *quantum* aggiuntivo di prudenza obiettivamente esigibile.

Qualora, invece, sia dato riscontrare, nei casi di volta in volta analizzati, l'assenza di percepibili "segnali d'allarme" circa i vizi gravi del *software*, ben potrà applicarsi –in favore del soggetto agente- il c.d. principio di affidamento, la cui attuazione è pienamente conforme con il principio di personalità della responsabilità penale ed altresì rispettoso dei doveri solidaristici gravanti sull'individuo che agisce nella consapevolezza del suo inserimento in una data formazione sociale al quale parametra la sua libertà d'agire.

Volendo riassumere quanto fin qui detto, rimanendo nello scenario del concorso colposo nel reato colposo altrui, connotato da un'azione sinergica dei diversi compartecipi (programmatore del *software* e autore materiale del reato presupposto 231), si potrebbe ragionevolmente inquadrare la responsabilità di entrambi:

(i) nel paradigma tipologico delle "violazioni cautelari concorrenti" che, pur afferenti a parametri comportamentali distinti, conducono entrambe alla verificazione dell'evento[56]; siamo in presenza di violazioni che investono parametri cautelari tra loro interferenti sul piano della gestione di un rischio comune. In questo caso – in ragione dell'alterità complementare che contraddistingue le attribuzioni di questi soggetti – è possibile intravedere due diversi illeciti monosoggettivi: quello "antecedente" del programmatore che per imperizia genera il *software,* destinato ad essere impiegato nella *compliance* aziendale, cui si innesta illecito colposo dell'autore materiale del reato presupposto 231. L'unitarietà delle due condotte si palesa nella connessione forte che le due violazioni instaurano sul piano del rischio: una connessione che, nel caso in esame, non vale certamente a determinare "nuove"

[56] G. MINICUCCI, *La plurisoggettività nell'agire colposo Una rilettura in chiave normativa. Una rilettura in chiave normativa,* op cit., pp. 183 ss. e 240 ss.

ISSN 2036-7821

incriminazioni – essendo il campo dominato dalla violazione di standard preventivi di primo grado – ma che consente l'applicazione dell'art. 113 c.p..
(ii) ovvero nel paradigma della "violazione cautelare realizzata dall'autore materiale del reato presupposto 231 e colposamente *indotta* dalla condotta involontaria di altro compartecipe" (nella specie il programmatore), secondo il modello dell'*autorìa mediata colposa*. Si è qui in presenza di una condotta negligente del compartecipe – pur non direttamente preventiva del rischio – la cui traduzione in pericolo si attualizza mediante il comportamento dell'autore "principale" dell'illecito colposo.
Il caso da noi analizzato presenta una notevole prossimità concettuale con l'art. 48 c.p., dal quale tuttavia diverge sotto un duplice profilo: da un lato, poiché l'induzione in errore è a sua volta involontaria; dall'altro, poiché la violazione dello standard di riferimento si riflette direttamente sul *quomodo* della condotta dell'autore materiale.
A fronte delle surriferite problematiche teoretiche offerte dagli istituti di parte generale, ragionando sul terreno prasseologico e in termini *de lege ferenda*, si potrebbero tipizzare e punire a titolo di colpa le seguenti esemplificative condotte dei programmatori, quali : i) l'omessa predisposizione da parte del programmatore di meccanismi preventivi nell'algoritmo che governa l'I.A.; ii) la mancata previsione di meccanismi di disattivazione del *software*, ovvero la disaccorta disattivazione di quei presidi di sicurezza di cui invece fosse stato originariamente dotato.
Ma qui verrebbe in rilievo un'altra problematica insita nella circostanza che spesso i programmi di intelligenza artificiale sono il frutto di un lavoro di squadra in cui cooperano un numero molto rilevante di tecnici e programmatori[57]. La "nascita" di un individuo artificiale intelligente richiede inevitabilmente la sinergia di competenze diversificate[58].
L'attribuzione della responsabilità ad una pluralità di coagenti implica infatti la soluzione di quello che la dottrina angloamericana chiama *many hands problem*; senza trascurare l'ipotesi in cui nella costruzione ed evoluzione di una forma di intelligenza artificiale potrebbe intervenire un'altra intelligenza artificiale, il che moltiplica i profili fattuali da considerare.

5.1 – La responsabilità dei c.d. "sorveglianti" del software.

Altro terreno di riflessione penalistica per individuare eventuali ulteriori responsabili della verificazione dei reati "presupposto 231" (consumati dai dipendenti o dai vertici dell'ente fruitore del software utilizzato per l'elaborazione dei protocolli

[57] La questione è stata autorevolmente sollevata da F. CONSULICH, op. loc. cit., p. 1040 ss.
[58] Si tratta di un problema che precede l'intervento penale e attiene alla stessa attribuibilità personale e giuridica del fatto; sul punto vedi D.J. GUNKEL, *Mind the Gap: Responsible Robotics and the Problem of Responsibility*, in *Ethics and Information Technology*, 2017, p. 12 ss..

ISSN 2036-7821

cautelari) potrebbe essere quello dei *sorveglianti*, i c.d. *human-in-the-loop*, ossia quei soggetti —individuati, di volta in volta, dalla normativa di settore e— *incaricati di supervisionare* il funzionamento del sistema di intelligenza artificiale, che viene poi impiegato nelle attività di *risk assessement* preliminari all'elaborazione del modello organizzativo. A tali soggetti non spetta solo il compito di vigilare sul funzionamento del dispositivo intelligente, ma anche quello di intervenire tempestivamente, qualora emergano malfunzionamenti.

La necessità che vi sia sempre un essere umano in grado di esercitare un controllo significativo sul funzionamento del sistema è prevista dalla gran parte delle fonti internazionali, sovranazionali e nazionali[59].

Tale principio dello *human oversight* è uno dei perni attorno ai quali ruota il recente Regolamento dell'U.E. sull'intelligenza artificiale (c.d. AI Act)[60], costituendo — nelle intenzioni del legislatore europeo — la modalità principe per conservare la centralità dell'essere umano in ogni fase dello sviluppo e dell'implementazione delle pratiche di intelligenza artificiale.

Occorre pertanto, ai fini della nostra ricerca, prendere le mosse dai contenuti del predetto Regolamento, che rappresenta il primo esempio, non soltanto a livello europeo ma globale, di normativa orizzontale — a carattere sistematico — sul terreno dell'intelligenza artificiale. Una delle sue caratteristiche distintive è infatti quella di travalicare i confini territoriali dell'Unione Europea per applicarsi, appunto

[59] Per un riepilogo delle disposizioni che, a livello statale o internazionale, richiedono la sorveglianza umana sui sistemi di I.A. si rinvia a B. G REEN, *The flaws of policies requiring human oversight of government algorithms*, in *Computer Law & Security Review*, 2022. p. 6 ss.

[60] Regolamento n. 2024/1689 sull'intelligenza artificiale (c.d. AI Act), pubblicato sulla Gazzetta Ufficiale dell'Unione Europea il 12 luglio 2024 ed entrato in vigore in data 1° agosto 2024. Tale Regolamento sull'intelligenza artificiale viene considerato, a livello globale, come il primo atto normativo volto a fornire un quadro normativo vincolante e orizzontale (ne senso che non è limitato ad un singolo settore di applicazione) per lo sviluppo e la commercializzazione di sistemi di I.A. La lettera a), dell'art. 2, 1° co., prescrive l'applicazione del regolamento "ai fornitori che immettono sul mercato o mettono in servizio sistemi di I.A. o immettono sul mercato modelli di IA per

finalità generali nell'Unione, indipendentemente dal fatto che siano stabiliti o ubicati nell'Unione o in un paese terzo". Rispetto alla versione originaria della proposta, l'attuale lett. a) contiene due elementi innovativi. Il primo è costituito dal riferimento ai modelli di IA per finalità generali, fattispecie che non era affatto contemplata nella proposta originaria della Commissione e che è stata successivamente integrata nel testo a seguito della grande diffusione dei c.d. *large language modelscome* GPT di Open A.I. L'aggiunta è pertanto coerente con la differenziazione del regime tra modelli di I.A. per finalità specifiche e per finalità generali. Il secondo elemento innovativo, di certo meno rilevante, è rappresentato dal criterio fattuale dell'"ubicazione" nell'Unione in aggiunta a quello tecnico-giuridico dello "stabilimento". Per effetto di queste modifiche, la lett. a) dispone l'applicabilità della normativa in oggetto a qualsiasi fornitore che immetta sul mercato o metta in servizio sistemi di I.A. (o modelli di I.A. per finalità generali) nell'Unione, indipendentemente dal fatto che tali fornitori siano stabiliti o ubicati all'interno o all'esterno del territorio dell'Unione. Per la trattazione esaustiva dell'argomento del perimetro territoriale applicativo del citato Regolamento, si rinvia ad un recente contributo di G. RESTA, *L'ambito territoriale di applicazione del Regolamento europeo sull'intelligenza artificiale: note critiche*, in *Il diritto dell'informazione e dell'informatica*, 2024, 6, pp. 731 – 745.

ISSN 2036-7821

orizzontalmente negli Stati membri[61], a qualsiasi modello di I.A., il cui utilizzo (o l'utilizzo dei relativi output) sia in grado di produrre effetti giuridicamente rilevanti nello spazio giuridico europeo[62].

L'art. 14 del citato Regolamento, rubricato "Sorveglianza umana" stabilisce che i sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso (par. 1). Nello specifico, il par. 4 dell'art. 14 elenca una serie di misure tecniche che devono essere predisposte dal fornitore, al fine di garantire un effettivo esercizio della sorveglianza "umana" sull'attività algoritmica (c.d. "misure di sorveglianza in senso stretto").

Due sono le tipologie di misure di sorveglianza umana, a seconda che si tratti di sistemi che necessitano di un filtro umano per incidere sul mondo (c.d. *recommendation systems*, ad esempio un *clinical decision support system*) o sistemi che, invece, incidono direttamente sulla realtà (ad esempio, un veicolo a guida autonoma, o un drone dotato di capacità di apprendimento).

Nel primo caso, la persona fisica deve essere messa in grado di decidere, in qualsiasi situazione particolare, di non usare il sistema di I.A. ad alto rischio o altrimenti di ignorare, annullare o ribaltare *l'output* del sistema di I.A. ad alto rischio (art. 14 par. 4, lett. d).

Nel secondo caso, invece, la persona fisica dovrà poter intervenire sul funzionamento del sistema di IA ad alto rischio o interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga che consenta al sistema di arrestarsi in condizioni di sicurezza (par. 4, lett. e).

Il potere di controllo attribuito al supervisore avrà, insomma, un carattere prettamente "negativo": non si richiede, infatti, che il sorvegliante debba essere messo nelle condizioni di svolgere, al posto del dispositivo, la funzione alla quale il dispositivo stesso era predisposto; si prevede, piuttosto, che questi possa arrestare il

---

[61] Al fine di adottare nell'ordinamento italiano le disposizioni del Regolamento UE 2024/1689 sull'intelligenza artificiale, al Senato (Atto n. 1146) è stato di recente approvato in prima lettura il *disegno di legge in materia di Intelligenza Artificiale;* in data 25 giugno 2025 il testo è stato approvato alla Camera (Atto Camera n. 2316) con modifiche e quindi il testo è tornato al Senato per una terza lettura. L'obiettivo del citato disegno di legge è la promozione di "un utilizzo corretto, trasparente e responsabile, in una dimensione antropocentrica, dell'intelligenza artificiale, volto a coglierne le opportunità" nonché l'introduzione di criteri regolatori, norme di principio e di settore, al fine di promuovere l'utilizzo delle nuove tecnologie, prevedendo, al contempo, misure in grado di contenere il rischio connesso al loro uso improprio o dannoso. I testi sono disponibili online su: *https://www.camera.it/leg19/126?&leg=19&idDocumento=2316*

[62] Il disegno di legge sull'intelligenza artificiale, licenziato dal Senato il 20 marzo 2025 e approvato con modificazioni dalla Camera il 25 giugno scorso, introduce in particolari aggravanti specifiche per i reati commessi mediante intelligenza artificiale (ad esempio in tema di aggiotaggio e manipolazione del mercato), nuove fattispecie autonome come la diffusione di contenuti *deepfake* (articolo 613-quater del Codice penale) e persino il reato di *scraping* (estrazione automatica di dati da siti web) o *data mining* (analisi di grandi quantità di dati per scoprire informazioni utili) abusivo su opere protette.

ISSN 2036-7821

sistema, o non prendere in considerazione la raccomandazione fornita dal sistema di I.A..

Nonostante l'art. 14 dell'AI-Act non imponga obblighi alla persona fisica alla quale è affidata la sorveglianza umana, si può facilmente immaginare che gli obblighi di progettazione posti in capo al fornitore troveranno un corrispettivo in appositi obblighi di controllo imposti alla persona incaricata della sorveglianza.

Al centro dell'AI-Act vi è proprio il concetto di rischio.

In particolare, l'AI Act divide i dispositivi di intelligenza artificiale in tre categorie — a seconda del rischio che questi pongono per la sicurezza degli utenti ed il rispetto dei diritti fondamentali dei cittadini —, individuando per ciascuna categoria un regime giuridico applicabile: così, ad esempio, i sistemi di I.A. che presentano *rischi considerati "inaccettabili"* non potranno essere introdotti nel mercato europeo (art. 5); i *sistemi ad alto rischio* (artt.6 e 7) potranno essere commercializzati, purché siano conformi rispetto ad una serie di requisiti (artt. 8-15); per i *sistemi a rischio limitato* (art. 50), infine, basterà il rispetto di alcuni obblighi di trasparenza.

Un ulteriore aspetto di rilievo introdotto dall'A.I. Act è la previsione di specifici obblighi non solo per fornitori, importatori e distributori di programmi di I.A., ma anche per i *deployer* (utilizzatori) degli stessi.

Ciò comporta, in un contesto di "*compliance* 231", la necessità di mappare accuratamente le responsabilità e i rischi-reato lungo l'intera filiera dei meccanismi algoritmici. L'A.I. Act potrebbe quindi rivelarsi un utile banco di prova per l'integrazione della *compliance* tecnologica con quella organizzativa e giuridica, favorendo l'aggiornamento dei modelli di cui al d.lgs. n. 231/2001 – con particolare attenzione al *risk assessment*, alla formazione sui rischi specifici dell'A.I. e agli *audit* periodici sui sistemi implementati –; e ciò sarà indispensabile non solo per evitare sanzioni, ma anche per accrescere la fiducia di *stakeholder* e mercati.

Il dovere di sorveglianza richiesto dal Regolamento europeo andrebbe poi affiancato e coordinato con quello dell'Organismo di vigilanza previsto dal decreto 231, che sarà chiamato a monitorare anche i rischi emergenti e a valutare l'effettività e l'efficacia delle misure adottate dall'impresa per mitigarli.[63]

L'operatore, lungi dall'essere una persona che adopera direttamente il dispositivo intelligente, rappresenta il soggetto che ha l'obbligo di *supervisionarlo* e di intervenire qualora emergano malfunzionamenti degli algoritmi che possano comportare situazioni di rischio o di offesa ai beni giuridici protetti (*rectius*: favorire o agevolare la commissione dei reati presupposto 231). In questa prospettiva, gli obblighi di sorveglianza potrebbero eventualmente assumere rilievo come obblighi di impedire determinati eventi lesivi (nella nostra ricerca: i reati annoverati nel catalogo 231),

[63] S. GUERRA, *L'impatto dell'Ai Act: presidi ad hoc per tracciare le decisioni dell'algoritmo. I modelli vanno integrati con procedure di verifica, audit e formazione,* in *NT-Plus Diritto*, 14 luglio 2025.

ISSN 2036-7821

integrando così la clausola di estensione della tipicità penale di cui all'art. 40, comma 2, c.p..

In siffatta ipotesi, ci troveremmo al cospetto di una duplice responsabilità a titolo di *cooperazione apparente*: da un canto, un fatto colposo o doloso -integrante il reato presupposto 231- dell'autore materiale (soggetto apicale o sub-apicale dell'ente), dall'altro, un'omissione impropria colposa del "complice" omittente (il c.d. sorvegliante o supervisore).

Il nodo principale resterebbe però il rapporto problematico tra diritto e intelligenza artificiale con le sue intrinseche peculiarità di autonomia, automazione e opacità dei meccanismi algoritmici, con il rischio di creare uno scollamento sistematico tra dovere di sorveglianza ed effettivo potere di sorveglianza[64]. E tale scollamento si accentua nel caso di sorveglianti di sistemi di I.A. dotati di automazione, ossia quei sistemi dotati di compiere una determinata attività, agendo direttamente sul mondo reale o virtuale in assenza di intervento umano[65].

L'alternativa al perimetro del cono applicativo dell'art. 40 cpv. c.p. (che presuppone l'esistenza di un garante), potrebbe essere quella, rimanendo sul terreno squisitamente dogmatico, di inquadrare la responsabilità dei sorveglianti (che hanno omesso di vigilare sul regolare funzionamento del *software* preposti alla predisposizione dei M.O.G.), nell'istituto di parte generale della *cooperazione colposa* ex art. 113 c.p,; responsabilità da declinarsi nel paradigma della *violazione cautelare dell'autore materiale, cui si correla un dovere preventivo disatteso a carico del compartecipe*[66] (ossia il sorvegliante). In quest'ultima evenienza, la genesi della situazione di rischio che ha favorito l'accadimento del reato presupposto 231 non promanerebbe dal contributo del compartecipe, bensì dall'autore materiale, protendendosi verso il titolare di doveri di controllo (il sorvegliante).

A fronte della paralisi dei meccanismi imputativi del diritto penale d'evento, e a fronte della problematica tematizzazione dogmatica di eventuali responsabilità

---

[64] B. FRAGASSO, *L'intelligenza artificiale e responsabilità penale,* op. cit., pp. 476 ss.

[65] B. PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'automazione artificiale*, in *Dir. inf.*, 2021, 1, p. 337 ss.; nonché A. CAPPELLINI, *Profili penalistici* delle self-driving cars, in *Dir. pen. cont.*, 2019, 2, p. 337.

[66] G. MINICUCCI, op.loc.cit., pp. 245 ss.; secondo l'Autore, nel caso di dovere preventivo disatteso dal compartecipe, il livellamento generale delle posizioni soggettive si pone in senso direttamente proporzionale alla sincronicità dell'agire ed inversamente proporzionale alla differenziazione delle competenze e, quanto ai suoi contenuti, passa per una iper-valorizzazione delle fonti di legge che descrivono responsabilità di organizzazione, obblighi di coordinamento o poteri di avocazione; le medesime fonti, peraltro, che sono del tutto contraddittoriamente evocate per responsabilizzare anche chi è destinatario di detti controlli, soprattutto nella parte in cui ritagliano spazi di autonomia valutativa che "imporrebbero" l'agire correttivo verso gli stessi soggetti sovraordinati. Sul punto, si veda altresì A. VALLINI, *Dimensioni sincroniche e diacroniche della colpa professionale*, in *Diritto e medicina: due scienze convergenti*?, a cura di G. LOSAPPIO, Taranto, 2019, p. 84 ss.. In questa ipotesi viene in gioco il mancato depotenziamento del pericolo in atto alla dimensione fisiologica del rischio consentito, ricorrendo doveri di controllo e di coordinamento che sfociano – o che, addirittura, devono sfociare – nell'impedimento dell'azione altrui.

ISSN 2036-7821

penali dei sorveglianti/supervisori, si potrebbero positivizzare talune fattispecie incriminatrici di pericolo "a tutela anticipata", che puniscano ad esempio la mancata realizzazione di una *rete protettiva* attorno all'operatività dell'I.A. ovvero l'omessa tempestiva attivazione di misure di sicurezza in caso di segnali di allarme[67], sempre che tali segnali siano riconoscibili (l'autonomia e la *black box* dei *software* possono infatti impedire al supervisore di riconoscere tempestivamente la sussistenza del pericolo che fa sorgere l'obbligo cautelare della ripresa del controllo).

Ad esse potrebbero affiancarsi altre ipotesi incriminatrici costruite sulla tecnica dell'inosservanza delle prescrizioni "legalmente date", per cui il privato (soggetto individuale) risponderebbe penalmente ove non ottemperi alla prescrizione di riparare una falla nei meccanismi preventivi di protezione individuata dalle competenti Autorità di controllo oppure non ripari prontamente i danni prodottisi a seguito di un disguido nel funzionamento dei meccanismi algoritmici. Bisognerebbe però "a monte" tipizzare protocolli cautelari rigidi, chiari e stringenti al fine di regolare l'interazione tra intelligenza artificiale ed "essere umano" secondo un paradigma di collaborazione "dinamica" tra il sorvegliante ed il sistema di I.A. che svolge la sua attività algoritmica.

Si tratta di assicurare costantemente la verificabilità – da parte della persona fisica "esperta"- delle scelte algoritmiche, minimizzando il rischio di errori o conseguenze imprevedibili, e garantendo all'individuo un pronto riacquisto del controllo della situazione in caso di necessità, per evitare o gestire rischi generati dall'I.A.. Difatti, il controllo della persona fisica sull'intelligenza artificiale può essere garantito solo tramite l'imposizione di presidi di progettazione e di monitoraggio delle azioni algoritmiche, non potendosi privare il sistema di I.A. di quel grado sufficiente di operatività e autonomia, che è tipico dei sistemi intelligenti dotati di automazione[68].

E proprio i rischi dell'imprevedibilità, dell'autonomia e dell'opacità (ove accertate) dei meccanismi algoritmici rappresenterebbero un argine invalicabile anche rispetto ai paradigmi ascrittivi della responsabilità penale del sorvegliante/supervisore, in particolare in punto di *esigibilità* di un obbligo di supervisione in capo a quest'ultimo.

Occorre, in definita, correttamente perimetrare un'area di rischio consentito, che trovi

[67] F. CONSULICH, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiale devianti*, in *Riv.it.dir.proc.pen.*, 2022, 3, p. 1051.

[68] Si rinvia alle proposte avanzate dalla *Singapore personal Data Protection Commission*, secondo la quale è necessario che, nei campi governati dall'intelligenza artificiale, sia sempre garantita la possibilità che la persona umana possa in ogni momento assumere nuovamente il controllo dello scenario operativo: *Singapore PDPC, A Proposed Model Artificial Intelligence Framework* , gennaio 2019, in *https://www.pdpc.gov.sg*. In particolare, si propone di implementare in tutti gli agenti artificiali sistemi simili alla scatola nera degli aeromobili, in modo da poter ricostruire i processi decisionali dell'AI. Si veda altresì il documento pubblicato dal Gruppo di esperti ad alto livello sull'intelligenza artificiale e indipendente, istituito dalla Commissione europea nel giugno 2018, intitolato *Etichs guidelines for Trustworthy Artificial Intelligence*, in *www.europa.eu*, 8 aprile 2019, p. 30 ss.; Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Creare fiducia nell'intelligenza artificiale antropocentrica*, 8 aprile 2019, *in eurlex.europa.eu*, p. 3 ss.

ISSN 2036-7821

la sua espressione nel principio di affidamento, non potendo privare il sorvegliante della facoltà di invocare a sua "discolpa" il fatto di aver confidato nella c.d. *diligenza algoritmica*[69], quanto meno in presenza di *software* dotati di automazione. Altrimenti il rischio è di trasformare il sorvegliante in un "capro espiatorio", a cui addebitare tutti gli eventi lesivi (*rectius*: ai fini della nostra indagine, i reati presupposto 231) derivanti dal sistema di I.A., anche nei casi in cui siano il frutto di un'attività imprevedibile e incontrollabile del dispositivo intelligente, fermo restando che anche l'affidamento alla diligenza algoritmica non può essere inteso alla stregua di una *delega in bianco all'algoritmo*.

6. Conclusioni.

Per il settore della *compliance* aziendale, la tecnologia rappresenta sia una sfida, che un'opportunità. Da un lato, aiuta a gestire le crescenti esigenze; dall'altro, però, comporta nuovi rischi.

L'intelligenza artificiale entra in gioco ogni volta che è necessario elaborare e analizzare grandi volumi di dati. Ciò che altrimenti richiederebbe ore e giorni, gli algoritmi intelligenti lo fanno in pochi secondi. Di conseguenza, possono ottimizzare i processi e alleviare il carico sui dipendenti umani.

Ciò apporta preziosi vantaggi anche in ambito *compliance.*

I sistemi di intelligenza artificiale possono aiutare a scoprire più rapidamente potenziali violazioni della *compliance* e irregolarità interne. Gli algoritmi sono in grado di identificare modelli noti e apprenderne di nuovi autonomamente.

L'intelligenza artificiale può aiutare nell'analisi e nella valutazione del rischio di compliance raccogliendo, valutando e correlando dati provenienti da molte fonti interne ed esterne.

L'intelligenza artificiale può filtrare documenti di *compliance* come regolamenti e standard, condurre un'analisi, dare priorità alle informazioni più rilevanti e far risparmiare ai/alle *Compliance Officer* molto tempo di lettura.

Ad esempio, un *chatbot* basato sull'intelligenza artificiale semplifica la comunicazione in modo comprensibile delle *policy* di compliance all'interno dell'azienda. Stiamo assistendo in questi ultimi anni ad una costante e progressiva digitalizzazione dei protocolli organizzativi e dei sistemi di analisi e di controllo dei rischi, adottati dagli enti allo scopo di tutelarsi dall'insorgere di responsabilità giuridiche.

È fuori discussione la progressiva influenza della I.A. rispetto alle modalità di controllo e di prevenzione adottate dalle società, talvolta, mediante integrazione degli strumenti "ordinari", talvolta, addirittura, in sostituzione di questi ultimi. Di qui, la neo-tematica della *Digital Criminal Compliance* nell'ambito della quale si

[69] In dottrina si è parlato di versione robotica del principio di affidamento, così B. FRAGASSO, *L'intelligenza artificiale e responsabilità penale,* op. cit., p. 503 e 507.

ISSN 2036-7821

trovano a dialogare due diversi "mondi", l'ordinamento giuridico e il settore informatico, con la conseguente trasposizione, in un contesto interdisciplinare, del tema della responsabilità amministrativa dell'ente[70].

Più nel dettaglio, l'innovazione tecnologica può consentire alle società – mediante, per esempio, l'utilizzo di specifici algoritmi – di migliorare la tempestività dei propri meccanismi di rilevazione delle irregolarità e/o delle condotte che potrebbero discostarsi o, addirittura, violare le disposizioni del Modello Organizzativo attuato dall'ente.

Gli algoritmi sono in grado di "captare" ed elaborare dati, in tempo reale, con conseguente maggiore probabilità di prevenire la violazione delle regole comportamentali e procedurali di cui al Modello 231.

Le nuove tecnologie, infatti, sono in grado di potenziare l'efficienza del sistema 231 con riferimento alla prevenzione, non solo, dei reati che si innestano in un tessuto informatico (ad esempi: i reati informatici, i cd. *market abuse*, i fenomeni di riciclaggio, ecc.) ma, più in generale, di tutte le fattispecie criminose "presupposto" che si possano verificare in seno all'ente.

In un ambito come quello dei modelli ex decreto 231/2001, dove la prevenzione dei rischi-reato rimane l'obiettivo principale, l'adozione di strumenti intelligenti ci pare irrinunciabile, alla stregua di una *best practice.*

Breve: il modello della responsabilità da reato degli enti si candida ad essere un paradigma essenziale nella gestione del rischio da «intelligenza artificiale»

E tuttavia non si può negare che l'I.A. manifesti taluni limiti e derive negative.

Infatti, in primo luogo, l'intelligenza artificiale può costituire un rischio di *compliance*.

Le decisioni prese da un'intelligenza artificiale dipendono sempre dai dati con cui è stata addestrata. Se il sistema operativo che genera gli algoritmi non è stato sufficientemente diversificato o riflette pregiudizi sociali, tali criticità si trasferiscono e si riverberano all'operatività degli algoritmi che governano un sistema di intelligenza artificiale.

Un algoritmo, infatti, può generare bilanci falsi, automatizzare richieste mendaci alla pubblica amministrazione; un sistema di intelligenza artificiale può alterare dati ambientali, ottimizzare l'elusione di limiti di emissione od orchestrare, in pochi millisecondi, una strategia di *pump and dump* o di *wash trading* su titoli quotati e, quindi, manipolare il mercato[71].

Come abbiamo osservato, un sistema informatico tradizionale viene sviluppato dal produttore e spesso viene personalizzato per lo specifico cliente. I risultati generati

[70] D. SPERANZINI, *Intelligenza artificiale e D.Lgs. 231/2001: i sistemi automatizzati di prevenzione dei reati*, in www.mmslex.com, 4 maggio 2022.

[71] Si rinvia in particolare a S. GUERRA, *Decreto 231, prevenire reati dell'intelligenza artificiale. Sono molte le violazioni che i sistemi di I.A. possono commettere,* in *NT-Plus Diritto*, 14 luglio 2025.

ISSN 2036-7821

dipendono dai seguenti fattori principali: la base di dati utilizzata, gli *input* ricevuti e il codice sorgente usato per svilupparlo.
A tali fattori principali, si aggiunge l'incidenza di altri due elementi ulteriori: la rete neurale, così come è stata definita dal produttore, e i dati con cui tale rete è stata addestrata. Prima di diventare operativa una I.A. ha bisogno di essere addestrata con enormi quantità di dati, che la plasmano e che determineranno le risposte fornite agli input. successivi. I dati con cui viene addestrata una I.A. possono essere forniti dall'utente o provenire da terzi e naturalmente devono essere di qualità per evitare un effetto ben noto in informatica: il cosiddetto GIGO, dall'inglese *Garbage In - Garbage Out* (spazzatura in entrata e spazzatura in uscita).
La valutazione di un sistema di IA non può limitarsi alla sola valutazione del cosiddetto codice sorgente della rete neurale, ma deve necessariamente considerare anche i dati che sono stati utilizzati per il suo addestramento, e gli effetti che questi hanno prodotto sulla rete neurale originale.
Quel che in sintesi si vuol dire è che affidarsi ciecamente all'intelligenza artificiale comporta taluni rischi, soprattutto quando le decisioni non possono più essere tracciate. Per questo si parla anche di effetto "scatola nera", poiché un'intelligenza artificiale apprende in modo indipendente, e cambia continuamente. Se il sistema operativo viene alimentato con dati errati o addirittura manipolato da hacker, può prendere decisioni sbagliate e causare malfunzionamenti. Ecco perché è importante controllare regolarmente il sistema di intelligenza artificiale, implementato nel contesto aziendale.
Vi è infatti che, in ambito di *Criminal Digital Compliance* – dove l'I.A. viene "applicata" al Sistema 231 –, gli algoritmi informatici sono chiamati a prevenire comportamenti devianti, che sono definiti tali alla stregua di disposizioni normative astratte (e mutevoli nel tempo) e che, in quanto tali, devono necessariamente essere oggetto di interpretazione.
La digitalizzazione del sistema di *criminal compliance* potrebbe presentare criticità anche sul piano pratico-operativo. Gli algoritmi presentano un limite (direi quasi connaturato): essendo meccanismi di previsione, strutturati su dati storici e finalizzati ad attivarsi allorché vengano rilevati elementi di allarme predeterminati, si corre il rischio che singoli o intere classi di individui siano considerati – dinanzi alla rilevazione di specifici elementi di rischio e, dunque, in via puramente prognostica ed ipotetica – "pericolosi" con riferimento alle modalità di svolgimento delle relative attività, con conseguente adozione della determinazione automatizzata.
Sicchè, l'ente, il quale intenda affidarsi a meccanismi digitalizzati di prevenzione dei reati, dovrà predisporre, contestualmente, adeguate tutele delle posizioni individuali che altrimenti rischiano di essere aprioristicamente e negativamente "attinte" da decisioni automatizzate distorte.

ISSN 2036-7821

In definitiva, non ci si può quindi limitare all'implementazione, *sic et simpliciter*, dei sistemi automatizzati nella *compliance* aziendale, ma tale implementazione dovrà essere accompagnata dalla predisposizione di una serie di misure ed accortezze – individuabili anche *dejure condendo*– idonee a prevenirne le conseguenti derive applicative scaturenti dalla operatività, necessariamente preventiva e prognostica, degli algoritmi.

Si può così pensare alla creazione e implementazione nella realtà aziendale di un nuovo protocollo cautelare (da inserire nella parte speciale dei M.O.G.); in dottrina[72] si è parlato efficacemente di "*Ai Governance Protocol*", da intendersi quale strumento che integra principi etici, presidi organizzativi e soprattutto talune cautele tecnologiche: quali, a titolo esemplificativo, registri degli algoritmi, *audit* tecnici, mappature del rischio tecnologico, sistemi di *flag* automatici sulle anomalie, separazione delle funzioni tra sviluppatori/programmatori e controllori, tracciabilità delle decisioni algoritmiche.

[72] S. GUERRA, op.loc.cit., 14 luglio 2025

ISSN 2036-7821