La responsabilità penale del produttore di sistemi di intelligenza artificiale [*]

di Beatrice Fragasso

Il contributo esamina i problemi ascrittivi della responsabilità penale in capo al produttore di sistemi di intelligenza artificiale (i.a.), per gli eventi lesivi derivanti dall'impiego di questi ultimi. L'obiettivo è di verificare se i tradizionali regimi di responsabilità penale – che hanno come presupposto la commissione di fatti penalmente rilevanti da parte di persone fisiche – siano adeguati al nuovo contesto tecnologico, in cui algoritmi caratterizzati da autonomia, interattività e opacità possono porre in essere condotte imprevedibili per gli stessi produttori. Da un lato, l'interposizione dell'imperscrutabile decision making *algoritmico tra la condotta del produttore e l'evento lesivo rende problematico l'accertamento del nesso di causalità, stante l'assenza, attualmente, di leggi scientifiche di copertura in grado di spiegare il comportamento dei dispositivi intelligenti. Dall'altro lato, la crisi del modello nomologico-deduttivo sembrerebbe ripercuotersi, a cascata, sull'accertamento della colpa in capo al produttore, al quale soltanto con gravi forzature potrebbe essere rimproverato il verificarsi di un evento lesivo concretamente imprevedibile. Preso atto delle criticità applicative del tradizionale diritto penale d'evento, il contributo vaglierà, infine, l'opportunità di fare ricorso a tecniche di anticipazione della tutela penale.*

SOMMARIO: 1. Premessa. – 1.1. Una nozione "penalisticamente orientata" di intelligenza artificiale. – 2. Il *machine learning*: una tecnologia *unpredictable by design*. – 3. La responsabilità civile del produttore (cenni). – 4. La responsabilità penale del produttore. – 5. L'accertamento del nesso di causalità: incompatibilità tra ragionamento causale e approccio probabilistico del *decision making* algoritmico? – 6. L'accertamento della colpa del produttore. – 6.1. Le regole cautelari scritte. – 6.2. Il rapporto tra regole cautelari scritte e regole cautelari non scritte: quale spazio per il rischio consentito? – 7. Una tutela anticipata in relazione ai sistemi di i.a. pericolosi: prospettive *de jure condito* e *de jure condendo*. – 8. Conclusioni.

1. Premessa.

Il cambiamento che l'intelligenza artificiale (d'ora in avanti, i.a.) sta portando nelle nostre vite è epocale. Già oggi, molti sistemi di i.a. hanno dimostrato di essere più accurati, nello svolgere le attività loro delegate, degli esseri umani: si pensi, ad esempio, all'attività di riconoscimento biometrico, che i sistemi di i.a. sono in grado di svolgere in

[*] Testo, rivisto e corredato di note, della relazione tenuta al Corso "Intelligenza artificiale, diritto e processo" organizzato dalla Scuola Superiore della Magistratura e dalla Fondazione Vittorio Occorsio (Napoli, 20-22 marzo 2023).

Editore Associazione "Progetto giustizia penale" | via Festa del Perdono 7, 20122 Milano
c/o Università degli Studi di Milano, Dipartimento di Scienze Giuridiche "Cesare Beccaria"
redazione@sistemapenale.it

modo più efficace (oltre che rapido) rispetto agli esseri umani[1]; si pensi, ancora, alle *self driving cars*, che, secondo alcune stime, se adottate da gran parte degli utenti della strada, potrebbero ridurre del 90% gli incidenti stradali[2] – oltre che liberare uomini e donne da un'attività spesso percepita come noiosa e stressante. Se la diffusione dell'intelligenza artificiale in ogni aspetto della vita quotidiana promette di arrecare grandi benefici alla società, essa, tuttavia, apre anche una serie di interrogativi sul piano politico, etico, economico, e, per quanto qui interessa, giuridico.

La questione, per quanto concerne i profili strettamente penalistici, può essere riassunta in poche battute: la (parziale) perdita di controllo dell'operatore umano (dell'utilizzatore, così come del produttore, del programmatore, dello sviluppatore, etc.) sul processo decisionale e sul comportamento dell'algoritmo potrebbe scardinare i classici meccanismi imputativi del diritto penale, comportando un'attenuazione, se non un totale annullamento, delle istanze punitive.

Chi risponde, allora, se un veicolo a guida autonoma attraversa un incrocio con il semaforo rosso e investe un pedone, cagionandone la morte?[3] Sono individuabili dei soggetti personalmente responsabili per le notizie false fornite da Chat GPT[4] e per le manipolazioni del mercato realizzate da algoritmi di *trading* finanziario?[5]

Il diritto penale – fondato sul *mancato dominio*, da parte dell'agente, di un *fatto offensivo effettivamente dominabile*[6] – rischia di risultare inadeguato laddove tale *dominio* si venga a perdere e "autore" immediato del reato risulti essere proprio una macchina.

Il problema del c.d. *responsibility gap*[7] è avvertito anche dalle istituzioni europee, che in diverse occasioni – riferendosi, in realtà, soprattutto a profili civilistici, di

---

[1] Vd. il report ID R&D, *Human or Machine: AI Proves Best at Spotting Biometric Attacks*, 2022.

[2] Vd. il report McKinsey & Company, *Ten ways autonomous driving could redefine the automotive world*, 1 giugno 2015.

[3] Ci sono diversi *database online* che raccolgono e classificano gli incidenti che hanno coinvolto veicoli a guida autonoma; tra i più aggiornati vd. il sito *Autonomous Vehicle Crashes* (https://www.avcrashes.net), che dà anche la possibilità di visualizzare i dati attraverso una mappa interattiva. In data 2 maggio 2023, il sito contava 630 incidenti, realizzatisi tra il 14 ottobre 2014 e il primo maggio 2023.

[4] Come noto, ChatGPT è un modello di i.a. "generativo", liberamente accessibile *online*, che, attraverso algoritmi di apprendimento automatico, genera risposte agli *input* scritti forniti dall'utente. Già in diversi casi è stata riscontrata la comunicazione di notizie false da parte del sistema, vd. D. Cassens Weiss, *ChatGPT falsely accuses law prof of sexual harassment; is libel suit possible?*, in *ABA Journal*, 6 aprile 2023; S. Khatsenkova – N. Huet, *Mayor mulls defamation lawsuit after ChatGPT falsely claims he was jailed for bribery*, in *Euronews.next*, 8 aprile 2023.

[5] Con l'espressione "*algorithmic trading*" ci si riferisce a sistemi informatici che sono in grado di effettuare autonomamente ordini di acquisto sulle piattaforme di *trading*: l'algoritmo decide la tempistica, il prezzo e la quantità dell'ordine; nella maggior parte dei casi, può addirittura avviare l'ordine in assenza di intervento umano. In diverse occasioni si sono già verificati fenomeni di improvvise turbative dei prezzi dei titoli finanziari, dovuti all'agire contestuale di "sciami" di algoritmi di *trading* (c.d. *Flash Crash*); in argomento vd. G. SCOPINO, *Algo Bots and the Law: Technology, Automation, and the Regulation of Futures and Other Derivatives*, Cambridge University Press, 2020; F. CONSULICH, *Il nastro di Möbius. intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca borsa*, 2018, vol. 71, n. 2, p. 195 ss.

[6] Così A. FIORELLA, *Responsabilità penale* (voce), in *Enc. Dir.*, vol XXXIX, 1988, p. 1289.

[7] Di *responsibility gap* si è parlato, inizialmente, soprattutto in dottrina. Vd., in particolare, i lavori seminali di A. MATTHIAS, *The responsibility gap: Ascribing responsibility for the actions of learning automata*, in *Ethics and Information Technology*, 2004, vol. 6, p. 175; R. SPARROW, *Killer robots*, in *J. Appl. Philos.*, 2007, vol. 24, p. 62; W.

responsabilità extracontrattuale – hanno evidenziato come i "comportamenti emergenti"[8] delle macchine "intelligenti" possano comportare un intollerabile vuoto di tutela nei confronti delle persone danneggiate. Già nel 2017, ad esempio, nella Risoluzione sul diritto civile e la robotica del Parlamento europeo (2015/2103(INL))[9], si poteva leggere, tra i considerando, che «più i robot sono autonomi, meno possono essere considerati come meri strumenti nelle mani di altri attori (quali il fabbricante, l'operatore, il proprietario, l'utilizzatore, ecc.)»[10] e che l'attuale quadro giuridico potrebbe rivelarsi non idoneo «a coprire i danni causati dalla nuova generazione di robot, in quanto questi possono essere dotati di capacità di adattamento e di apprendimento che implicano un certo grado di imprevedibilità nel loro comportamento, dato che imparerebbero in modo autonomo, in base alle esperienze diversificate di ciascuno, e interagirebbero con l'ambiente in modo unico e imprevedibile»[11].

Posto che l'ipotesi – pur avanzata in dottrina – di una responsabilità penale diretta del dispositivo intelligente[12] non sembra condivisibile, a meno di non voler trasfigurare completamente i connotati del diritto penale[13], ci pare che la riflessione sui

---

WALLACH – C. ALLEN, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2009, p. 198 ss. Recentemente si vedano le acute riflessioni di A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, in G. Balbi e al. (a cura di), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"*, Giappichelli, 2023, *passim*.

[8] Sul concetto di "*emergent behaviors*" – utilizzato in letteratura per descrivere i comportamenti "autonomi" dei sistemi di i.a. – vd., per tutti, R. CALO, *Robotics and the Lessons of Cyberlaw*, in 103 *Calif. L. Rev*., 2015, *passim*, spec. p. 532 ss.

[9] Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

[10] *Ibidem*, considerando AB.

[11] *Ibidem*, considerando AI.

[12] Il principale teorizzatore della possibile configurazione di una responsabilità penale diretta in capo ai sistemi di i.a. è il penalista israeliano Gabriel Hallevy, vd. G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; ID., *"I, Robot – I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offenses*, in 22 *Syracuse Sci. & Tech. L. Rep. 1*, 2010, p. 1 ss.; ID., *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in 4 *Akron Intellectual Property Journal*, 2010, p. 171 ss.; nello stesso senso, ma con accezioni spesso molto diverse tra loro, vd. Y. HU, *Robot Criminals*, in 52 *U. Mich. J. L. Reform*, 2019, p. 487 ss.; C. MULLIGAN, *Revenge Against Robots*, in 69 *South Carolina Law Review*, 2018, p. 579; M. SIMMLER – N. MARKWALDER, *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Crim. Law Forum*, 2019, n. 30, p. 1 ss.; F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective*, in *Philosophy & Technology*, 2020, 33, p. 433 ss.

[13] Non è questa la sede per cercare di confutare, passaggio per passaggio, le argomentazioni proposte dai fautori della tesi della responsabilità penale diretta dei sistemi di i.a. Ci limitiamo qui a sottolineare che il principio di *personalità della responsabilità penale* (art. 27, co. 1, Cost.) pare un ostacolo insormontabile al riconoscimento di una responsabilità penale diretta del sistema intelligente, presupponendo una capacità di autodeterminazione dell'agente che, ad oggi, caratterizza esclusivamente gli esseri umani (seppure, anche in questo caso, soltanto come postulato epistemologico). Del pari, anche a voler riconoscere una "capacità criminale" in capo ai sistemi di i.a., non si vede quali sanzioni, applicate agli algoritmi, possano rispondere alle finalità classiche della pena (retribuzione; prevenzione generale e speciale). Per un'approfondita confutazione della tesi della responsabilità penale diretta delle macchine vd., per tutti, A. CAPPELLINI, Machina delinquere non potest? *Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*,

reati da intelligenza artificiale dovrebbe coinvolgere, da un lato, la sfera dell'utilizzatore del sistema di i.a., e, dall'altro, quella del "produttore". In questa sede, nell'attenerci al tema assegnatoci dagli organizzatori del Corso, ci concentreremo esclusivamente su quest'ultimo aspetto.

Premettiamo fin da ora che, per comodità espositiva, in questa relazione si farà sempre riferimento alla figura del "produttore", intendendo tuttavia includere, in tale espressione, tutte quelle persone che, a vario titolo, contribuiscono ai processi di *sviluppo, progettazione e commercializzazione* dei dispositivi intelligenti. Va da sé, ovviamente, che l'individuazione del soggetto personalmente responsabile – tra coloro che fanno parte del ciclo *lato sensu* produttivo – incontrerà difficoltà specifiche, determinate dal problematico accertamento: (*i*) della specifica *causa* dell'evento lesivo (es. difetto di programmazione o di addestramento o di installazione, etc.); (*ii*) della persona responsabile all'interno delle organizzazioni complesse.

1.1. La necessità di una nozione "penalisticamente orientata" di intelligenza artificiale.

Nell'ambito di questo Corso sono già ampiamente emerse le complessità del rapporto tra sistema penale e intelligenza artificiale. Concludiamo queste giornate con molti spunti di riflessione, ma quasi nessuna certezza – se non, forse, con una sola: che non esistono, ad oggi, definizioni unanimemente condivise di i.a. L'hanno sottolineato tutti i relatori e anche questo intervento partirà da qui.

Una definizione destinata ad avere importanti ripercussioni è sicuramente quella fornita dalla proposta di Regolamento sull'intelligenza artificiale presentata nel 2021 dalla Commissione Europea (c.d. *AI Act*)[14], che definisce il sistema di i.a. come «un *software* sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare *output* quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono» (art. 3, lett. a)[15]. Si tratta, com'è evidente, di una definizione molto ampia, e che per questo motivo è stata molto criticata in dottrina, poiché estende il campo di

---

2018, pubblicato successivamente in *disCrimen*, 2019, p. 14 ss.; C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 2020, n. 4, p. 1766; B. PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, 2021, fasc. 1, p. 345 ss.

[14] Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, COM/2021/206 final, 21 aprile 2021 (cd. *AI Act*)

[15] Gli Approcci indicati dall'allegato 1 sono: «a) Approcci di apprendimento automatico, compresi l'apprendimento supervisionato, l'apprendimento non supervisionato e l'apprendimento per rinforzo, con utilizzo di un'ampia gamma di metodi, tra cui l'apprendimento profondo (*deep learning*); b) approcci basati sulla logica e approcci basati sulla conoscenza, compresi la rappresentazione della conoscenza, la programmazione induttiva (logica), le basi di conoscenze, i motori inferenziali e deduttivi, il ragionamento (simbolico) e i sistemi esperti; c) approcci statistici, stima bayesiana, metodi di ricerca e ottimizzazione».

applicazione della proposta anche a sistemi che normalmente, oggi, non sono considerati come "intelligenti" e che spesso sono già in circolazione da decenni[16].

Per il tema che qui ci proponiamo di indagare, emerge piuttosto la necessità di individuare una nozione "penalisticamente orientata" di intelligenza artificiale, che comprenda soltanto i sistemi che hanno caratteristiche *dirompenti* rispetto ai classici modelli di imputazione della responsabilità penale. Dovrebbero dunque essere esclusi dalle nostre prospettive di studio tutti quei sistemi che – essendo inquadrabili all'interno dei tradizionali schemi di *strumentalità* dell'oggetto rispetto all'agente umano – riproducono problematiche giuridiche note.

Il *cuore del problema* sembrerebbe consistere nell'*imprevedibilità* del comportamento di alcuni dispositivi intelligenti – caratteristica che, a sua volta, parrebbe derivare da tre proprietà che si riscontrano nei più sofisticati sistemi di i.a., l'*autonomia*, l'*interattività*, l'*opacità*:

(i) *autonomia* – è la caratteristica, evidenziata dalla gran parte della dottrina che si è occupata del tema[17], che rischia di scardinare i meccanismi classici di imputazione della responsabilità penale. Si tratta di un concetto che dev'essere inteso in senso restrittivo: non nell'accezione kantiana di *capacità morale di auto-governo della ragione*, ma, piuttosto, come *capacità di prendere decisioni in situazioni di incertezza* e di *compensare l'incompletezza delle informazioni ricevute in partenza attraverso l'apprendimento*;

(ii) *interattività* – gli algoritmi intelligenti spesso non agiscono come monadi, ma sono connessi tra loro e, talvolta, persino con l'ambiente fisico in cui si trovano (c.d. *internet of things*[18]). Da un lato, l'incontro tra due o più "autonomie" – nell'accezione anzidetta – può comportare *output* collettivi inaspettati (c.d. *collective machine behaviour*[19]), frutto di interazioni non sempre pienamente comprensibili dall'esterno; dall'altro, gli algoritmi connessi con sensori e infrastrutture fisiche complesse sono esposti ad una varietà infinita di *inputs*, risultando così non meno imprevedibili dell'ambiente con cui interagiscono[20];

---

[16] C. OSBORNE, *The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem*, in *Centre for Data Ethics and Innovation Blog*, 11 maggio 2021; L. Clarke, *The EU's leaked AI regulation is ambitious but disappointingly vague*, in *Tech Monitor*, 15 aprile 2021. Alcune fonti, in ogni caso, riportano che Consiglio e Commissione Europea si stiano accordando per restringere la definizione e limitare il campo di applicazione del Regolamento ai soli sistemi dotati di *machine learning*, vd. N. Kayser-Bril, *European Council and Commission in agreement to narrow the scope of the AI Act*, in *Algorithm Watch*, 24 novembre 2021.

[17] Vd., per tutti, L. PICOTTI, *I primi vent'anni della Convenzione di Budapest nell'ottica sostanzialista e la mancata ratifica ed esecuzione del Primo Protocollo addizionale contro il razzismo e la xenofobia*, in *Dir. pen. proc.*, 2022, n. 8, p. 1032; A. AMIDEI, *Intelligenza Artificiale e* product liability*: sviluppi del diritto dell'Unione Europea*, in *Giur. it.*, luglio 2019, p. 1717.

[18] Non esiste una definizione univoca di *internet of things*. L'ENISA (*European Union Agency for Cybersecurity*) lo definisce come «*a cyber- physical ecosystem of interconnected sensors and actuators, which enable intelligent decision making*»; lo European Research Cluster on the Internet of Things (IERC) lo definisce invece come «*a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network*».

[19] I. RAHWAN E AL., *Machine Behaviour*, in *Nature*, 2019, n. 568, p. 482.

[20] C.E.A. KARNOW, *The application of traditional tort law*, in R. Calo e al. (eds.), *Robot Law*, Edward Elgar Publishing, 2016, p. 59; In generale, sull'interattività dei sistemi di i.a. vd. A. BECKERS – G. TEUBNER, *Three*

(iii) *opacità* (c.d. *black box*) – gli algoritmi intelligenti non sono in grado di fornire una spiegazione teorica e causale dei risultati raggiunti, né i programmatori possono agevolmente intuirla. Per *black box* si intende l'imperscrutabilità dei meccanismi causali interni ai sistemi di i.a.: è possibile individuare *input* e *output*, ma non è invece possibile ricostruire *cosa accade* all'interno della scatola nera, ovverosia la catena causale che ha determinato il passaggio dagli *input* agli *output*[21]. Di *black box*, com'è noto, si parlava già con riferimento ai prodotti tradizionali[22]: in relazione ai più sofisticati sistemi di i.a., tuttavia, l'opacità sembrerebbe una caratteristica *intrinseca* e *ineliminabile*, derivante dalla discrepanza tra processo logico-computazionale probabilistico tipico dell'algoritmo e struttura causale e deduttiva propria del ragionamento umano.

I sistemi di i.a. che hanno le caratteristiche così descritte possono sostanzialmente essere ricondotti al *machine learning* – una tecnica, su cui torneremo a breve, che consente agli algoritmi di *apprendere* dall'esperienza e dall'ambiente circostante, modificando le proprie prestazioni nel corso del tempo[23]. Ci limitiamo qui a sottolineare che anche il Consiglio di Stato, in una pronuncia del 2021[24], ha adottato, seppur soltanto in un *obiter dictum*, una definizione restrittiva di intelligenza artificiale, limitata alle sole applicazioni di *machine learning*: nelle parole dei giudici amministrativi, l'i.a. è «un sistema che non si limita solo ad applicare le regole software e i parametri preimpostati (come fa invece l'algoritmo "tradizionale") ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico».

2. I sistemi di *machine learning*: una tecnologia *unpredictable by design*.

Senza entrare in aspetti tecnici troppo specifici – che di certo non abbiamo le competenze per fornire e che sono già stati accuratamente delineati nelle sessioni precedenti –, ci limiteremo qui ad evidenziare brevemente alcune delle caratteristiche dei sistemi di *machine learning* che ne rendono problematico l'inquadramento all'interno delle categorie del diritto penale.

Liability Regimes for Artificial Intelligence, Bloomsbury, 2022, p. 111 ss.

[21] Y. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in 31 *Harvard Journal of Law & Technology*, 2018, n. 2, p. 905; D. CASTELVECCHI, *Can We Open the Black Box of AI?*, in *Nature*, 5 ottobre 2016.

[22] Vd. per tutti F. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffrè, III ed., 2003, p. 224-235; C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Giuffrè, 2004, p. 50 ss.

[23] S. RUSSELL – P. NORVIG, *Artificial Intelligence: A Modern Approach*, Pearson College Div., 4th ed., 2020, p. 651 ss.

[24] Cons. Stato, sez III, sent. 25 novembre 2021, n. 7891; in argomento vd. i commenti di F. PAOLUCCI, *Algoritmi e intelligenza artificiale alla ricerca di una definizione: l'esegesi del Consiglio di Stato, alla luce dell'AI Act*, in *Quest. giust.*, 8 aprile 2022; C. FILICETTI, *Sulla definizione di algoritmo (nota a Consiglio di Stato, Sezione Terza, 25 novembre 2021, n. 7891)*, in *Giust. ins.*, 8 febbraio 2023

A differenza dei modelli simbolico-deduttivi di i.a. sviluppati a partire dalla metà del Novecento – che applicavano al caso concreto le regole astratte fornite in sede di programmazione[25] –, i sistemi di *machine learning* utilizzano un metodo induttivo (c.d. *bottom-up*) e probabilistico: osservano i dati forniti in sede di addestramento, riconoscono i *pattern* statistici sottostanti e ne traggono delle generalizzazioni[26].

Tale approccio parte dalla constatazione che molte delle attività che l'essere umano svolge non sono formalizzabili attraverso l'esplicitazione di regole definite, ma sono frutto di esperienza ed imitazione. Si pensi, ad esempio, al riconoscimento facciale: un'attività che svolgiamo senza difficoltà, in maniera intuitiva, ma per la quale sarebbe arduo stilare un elenco di regole precise che possano essere applicate, in via deduttiva, da una macchina. Se invece forniamo ad un algoritmo di *machine learning* una serie di immagini, indicando appositamente, attraverso un'etichetta (*label*), quali ritraggono Tizio e quali ritraggono Caio, il sistema ragionerà per analogia e, quando posto di fronte ad immagini non etichettate, sarà in grado di generalizzare quanto appreso in fase di addestramento, distinguendo il volto di Tizio da quello di Caio (c.d. apprendimento supervisionato o *supervised learning*)[27].

Un simile risultato sarebbe difficilmente raggiungibile attraverso l'approccio simbolico e "causale", che richiederebbe la definizione di un procedimento logico astratto, da applicare nel caso concreto. D'altra parte, lo stesso algoritmo non è in grado di formulare una *regola di classificazione* – che consenta *in tutti i casi*, in maniera infallibile, di etichettare le immagini – ma ragiona *per analogia*: il sistema non sa spiegare quali sono stati i criteri che, in un *database* contenente milioni di immagini, gli hanno consentito di individuare il viso di una determinata persona.

Così, se, da un lato, il *machine learning* consente agli algoritmi di individuare *pattern* ricorrenti che potrebbero non essere percepibili dall'uomo, dall'altro, il ragionamento analogico tipico dei sistemi di auto-apprendimento sconta una carenza assoluta di comprensione "semantica", che talvolta può determinare la commissione di errori che, ad uno sguardo umano, possono apparire come grossolani e macroscopici[28].

---

[25] I modelli simbolico-deduttivi di i.a. potevano applicare regole di tipo consequenziale (*if-this-then-that rules*), semplificando e velocizzando le attività umane, ma non erano in grado di gestire situazioni di incertezza. Questi modelli – sebbene utilizzati fino a non molto tempo fa – costituiscono quella che potremmo chiamare l'*archeologia* dell'intelligenza artificiale e sono oggi comunemente chiamati, con un misto di derisione e nostalgia, "*Good Old-Fashioned AI*" (GOFAI), vd. W. WALLACH – C. ALLEN, *Moral Machines: Teaching Robots Right from Wrong*, cit., p. 183; M.A. LEMLEY – B. CASEY, *Remedies for Robots*, in 86 *University of Chicago Law Review*, 2019, p. 1322-1323.

[26] Vd. per tutti S. RUSSELL – P. NORVIG, *Artificial Intelligence: A Modern Approach*, cit., p. 651; P. DOMINGOS, *A Few Useful Things to Know about Machine Learning*, in *Communications of the ACM*, vol. 55, n. 10, nov. 2012, p. 79.

[27] G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, il Mulino, 2022, cit., p. 50-51.

[28] In argomento vd. A.D. SELBST, *Negligence and AI's Human Users*, in 100 *Boston University Law Review*, 2020, p. 1318 ss. Sull'incapacità di comprensione simbolica dei sistemi di *machine learning* vd. per tutti M.B. MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen.*, 10 maggio 2020, p. 12.

Può capitare, allora, che un sistema di riconoscimento di immagini basato sul *machine learning* scambi una tartaruga per un fucile[29]. Ancora, alcuni ricercatori hanno dimostrato che possono bastare degli adesivi o dei graffiti su un segnale stradale per causare un errore di riconoscimento da parte di un sistema di guida autonoma[30]. Piccolissime variazioni nell'*input* – insignificanti agli occhi di un osservatore umano – possono determinare una percezione erronea da parte del sistema di i.a., proprio per il fatto che le tecniche di *machine learning* sono in grado di individuare associazioni statistiche nei dati, ma non sono invece capaci di tracciare modelli astratti di spiegazione causale di tali ricorrenze[31].

Per concludere questa introduzione di carattere fenomenologico, un aspetto che è fondamentale rimarcare è che il funzionamento delle tecniche più sofisticate di *machine learning* è ancora ignoto, nonostante se ne sfruttino ampiamente le potenzialità applicative: insomma, non si sa bene *perché*, ma le previsioni effettuate dagli algoritmi di i.a. sono, nella maggioranza dei casi, corrette, e spesso più accurate di quelle umane. Ci si affida ad esse come ad un *oracolo* o ad uno "stregone", tanto che i riferimenti alla *magia* e al *genio* sono frequenti da parte degli stessi ricercatori che sviluppano i sistemi di i.a.[32]

Con la diffusione dei sistemi di i.a., si assiste dunque ad una nuova fase del rapporto tormentato tra tecnologia e *agency* umana. In questo caso, infatti, l'imprevedibilità non è un *bug* nel sistema, ma piuttosto un risultato voluto e ricercato dagli stessi programmatori, poiché consente di raggiungere i risultati più efficienti: in questo senso, si dice che i sistemi di i.a. sono *unpredictable by design*[33].

---

[29] A. ATHALYE E AL., *Synthesizing Robust Adversarial Examples*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[30] K. EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[31] A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, in 87 *Fordham Law Review*, 2018, p. 1097; M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, in 26 *Science and Engineering Ethics*, 2020, p. 2060; Z.C. LIPTON, *The Mythos of Model Interpretability*, in 16 *Queue*, May-June 2018, p. 8-9.

[32] Così, ad esempio, afferma Pedro Domingos, uno dei più noti ricercatori in materia di *machine learning*: «*developing successful machine learning applications requires a substantial amount of 'black art' that is difficult to find in textbooks*», vd. P. DOMINGOS, *A Few Useful Things to Know about Machine Learning*, cit., p. 78; vd. anche R. Giacconi, *Una sorta di magia. Intelligenze artificiali, origami, marionette, serpenti: scoprire l'incanto nei laboratori di robotica*, in *Il Tascabile*, 2 marzo 2023: in cui si riportano le considerazioni di Jamie Paik, direttrice del Reconfigurable Robotics Lab (RRL) di Losanna: «ogni volta che accendiamo un robot, gli diamo vita e fa quello che ci aspettavamo, è un momento magico, di gioia, euforia. È come se dicessimo sempre una piccola preghiera: *fa' che funzioni questa volta, fa' che funzioni questa volta…*».

[33] R. CALO, *Robotics and the Lessons of Cyberlaw*, in 103 *Calif. L. Rev.*, 2015, p. 542; J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility: The prospect of expert robots*, in R. Calo e al. (eds.), *Robot Law*, cit., p. 107; vd. anche le riflessioni pioneristiche svolte, quando ancora non erano state sviluppate le tecniche di *machine learning*, da C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, in 11 *Berkeley Technol. Law J.*, 1996, p. 192.

3. La responsabilità civile del produttore di sistemi di i.a. (cenni).

Date queste brevi premesse fenomenologiche, ci concentreremo ora sugli aspetti più strettamente attinenti all'attribuzione della responsabilità per eventi lesivi derivanti dai sistemi di i.a.

Innanzitutto, può essere utile sottolineare che, ad oggi, il dibattito in sede normativa e accademica si è perlopiù focalizzato sui meccanismi ascrittivi della responsabilità *civile* del produttore[34]. La responsabilità aquiliana si candida infatti ad essere lo strumento cardine per la tutela delle persone danneggiate dai sistemi di i.a.: da un lato, grazie al ricorso al modello di imputazione oggettiva previsto dalla direttiva sul danno da prodotto; dall'altro, attraverso la predisposizione di meccanismi di agevolazione probatoria per il ricorrente.

Il testo fondamentale è ovviamente la direttiva sulla responsabilità per danno da prodotto (dir. 85/374/CEE[35]), che prevede la responsabilità oggettiva del produttore per i danni cagionati dal prodotto difettoso. La Commissione Europea, nel settembre 2022, ha presentato una proposta di direttiva volta ad aggiornare e a rendere applicabile tale testo normativo ai nuovi prodotti intelligenti e volta altresì ad introdurre delle inversioni dell'onere della prova a favore del danneggiato (in particolare, delle presunzioni di difettosità del prodotto e di causalità tra difetto e danno)[36]. Complessivamente, questi accorgimenti dovrebbero agevolare il ristoro della persona che ha subito danni provocati da sistemi di i.a. Cercando di riassumere al massimo, la strategia che sta cercando di implementare la Commissione europea parrebbe quella di *accollare economicamente al produttore il rischio dell'imprevedibilità del sistema di i.a.*, garantendo così, al consumatore, un rimedio risarcitorio efficace e facilmente esperibile[37].

---

[34] Nella letteratura italiana si vd., per tutti, il volume curato da A Pajno e al., *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1 e 2, Il Mulino, 2022, in particolare i contributi di U. RUFFOLO – A. AMIDEI, *La regolazione* ex ante *dell'intelligenza artificiale tra gestione del rischio* by design*, strumenti di certificazione preventive e «autodisciplina» di settore*; U. RUFFOLO, Artificial Intelligence *e responsabilità. «Persona elettronica» e teoria dell'illecito*; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione* by design *e obblighi di trasparenza*; vd. anche G. ALPA, *Quale modello normativo europeo per l'intelligenza artificiale?*, in *Contr. impr.*, 2021, n. 4, p. 1003 ss.; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata? Riflessioni a margine del percorso*, in *Nuova giur. civ. comm.*, 2020, n. 6, p. 1346; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, in *Oss. dir. civ. comm.*, 2020, fasc. 2, p. 445 ss.; in ambito europeo vd. A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., e i contributi contenuti in M. Ebers – S. Navas (eds.), *Algorithms and Law*, Cambridge University Press, 2020, e S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019.

[35] Direttiva 85/374/CEE del Consiglio del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi.

[36] Commissione europea, *Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi*, COM (2022) 495 final, 28 settembre 2022.

[37] La Commissione europea ha altresì presentato una Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), COM (2022) 496 final, 28 settembre 2022, che mira ad introdurre, per i danni cagionati da sistemi di i.a. ai quali non sia applicabile la Direttiva sulla responsabilità da prodotto, una *presunzione del nesso di causalità in caso di colpa* (art. 4) e la

4. La responsabilità penale del produttore di sistemi di i.a.

Se la responsabilità aquiliana si conferma uno strumento duttile, capace di adattarsi ai nuovi fenomeni tecnologici, di certo lo stesso non può dirsi dell'ordinamento penale. Le garanzie che circondano il sistema costituzionale del diritto penale (prime tra tutte: legalità, offensività, colpevolezza, presunzione di innocenza) lo rendono infatti un diritto *rigido*, inidoneo a contenere quei fenomeni sistemici, globalizzati ed incerti che caratterizzano la post modernità.

In particolare, il reato "da intelligenza artificiale" costituisce solo l'ultima tappa di quella tendenziale crisi del delitto d'evento che caratterizza, ormai da decenni, il diritto penale di fronte alla c.d. "società del rischio"[38]: è il noto "shock da modernità" di cui parlava Federico Stella[39].

La problematica riconducibilità del fatto algoritmico alla "persona" del produttore si innesta infatti in un contesto, già esistente, di grave messa in discussione (fino, talvolta, alla forzatura) delle categorie dogmatiche tradizionali, incapaci di contenere – se non attraverso notevoli flessibilizzazioni – la complessità dei fenomeni moderni. Nello specifico, il danno da dispositivo intelligente ripropone, accentuandoli, alcuni dei profili più problematici già sorti in relazione alla responsabilità penale per danno da prodotto: l'indebita sovrapposizione tra struttura commissiva e omissiva del reato; l'ostica identificazione dei soggetti personalmente responsabili all'interno delle organizzazioni complesse; l'individuazione del nesso di causalità in relazione a prodotti caratterizzati da opacità; l'accertamento della colpa in situazioni di incertezza scientifica.

Se le prime tra le questioni citate riproducono sostanzialmente, aggiornandoli, gli interrogativi già emersi in passato, l'imprevedibilità dei sistemi di i.a. porta invece ad un ulteriore livello di complessità l'accertamento del nesso eziologico tra condotta umana ed evento lesivo, nonché la valutazione della colpa in capo all'agente – con

---

possibilità, per l'autorità giudiziaria, di *ordinare la divulgazione degli elementi di prova*, qualora emergano indici di fondatezza della domanda risarcitoria (art. 3).

[38] Vd. C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Riv. it. dir. proc. pen.*, 2005, p. 1684; F. STELLA, *Giustizia e modernità*, cit.; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Leg. pen.*, 2011, n. 2, p. 397 ss.; J.M. SILVA SÁNCHEZ, *L' espansione del diritto penale. Aspetti della politica criminale nelle società postindustriali*, Giuffrè, 2004 (ed. spagn. 1999); F. HERZOG, *Società del rischio, diritto penale del rischio, regolazione del rischio*, in L. Stortoni – L. Foffani (a cura di), *Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte*, Giuffrè, 2004, p. 357; F. CENTONZE, *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*, Giuffrè, 2004. Il concetto di "società del rischio", come noto, si deve a U. BECK, *La società del rischio. Verso una seconda modernità*, Carocci, 2000 (ed. ted. 1986), che ha notevolmente influenzato le citate riflessioni penalistiche, ponendo l'accento su come le tecniche di produzione moderna creino sistematicamente rischi che la società, nel suo complesso, non è in grado di controllare ed assorbire.

[39] F. STELLA, *Giustizia e modernità*, cit., *passim*. Vd. sul punto F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, fasc. 10, p. 4.

particolare riferimento al requisito della *prevedibilità* dell'evento lesivo. È su questi aspetti ci concentreremo ora.

5. L'accertamento del nesso di causalità: incompatibilità tra ragionamento causale e approccio probabilistico del *decision making* algoritmico?

L'ostacolo maggiore all'accertamento del nesso di causalità, ad oggi, è costituito dall'assenza di leggi scientifiche consolidate che siano in grado di descrivere il dispiegarsi della catena causale nel funzionamento degli algoritmi di *machine learning*. Come è stato già accennato parlando della *black box*, in relazione agli algoritmi intelligenti è astrattamente possibile individuare *input* e *output*, ma non è invece agevole ricostruire la catena causale che lega un determinato *input* ad un determinato *output*, a causa del modello di ragionamento *probabilistico* che caratterizza il funzionamento dei sistemi di *machine learning*.

Secondo alcuni studiosi, a tale *gap* conoscitivo potrebbe sopperire l'installazione di *event data recorders* (c.d. EDR, comunemente conosciuti come "scatole nere" o "*logs*"), funzionali ad illuminare, *ex post*, la dinamica dell'evento lesivo[40]. Tali strumenti, tuttavia, non sembrano pienamente in grado di rimediare al *deficit* di *comprensibilità* delle decisioni algoritmiche. Per quanto riguarda gli autoveicoli, ad esempio, l'*event data recorder* può, a seconda delle versioni, registrare la velocità di crociera, l'accelerazione, gli eventuali *input* da parte del conducente (es. la pressione sul pedale del freno), l'attivazione di spie o segnali da parte del veicolo[41]. La scatola nera, tuttavia, *non può dire nulla sulle cause dell'incidente*: si limita a fornire *dati grezzi* da interpretare in sede processuale – attraverso l'ausilio di periti e consulenti tecnici – ma non è di per sé risolutivo ai fini della ricostruzione del nesso causale. Insomma, le questioni restano aperte: perché il veicolo a guida autonoma non ha rallentato di fronte al semaforo rosso? Perché non ha applicato lo spazio di frenata previsto? Il prodotto algoritmico era difettoso? E, se sì, quale componente difettosa del prodotto ha cagionato l'evento lesivo?

Può tra l'altro ipotizzarsi che, in futuro, diventerà tecnicamente possibile ricostruire *ex post* la catena causale che ha condotto alla verificazione di singoli eventi lesivi algoritmici – magari proprio attraverso lo sviluppo di sistemi di i.a. di *reverse*

---

[40] Così C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1760; G. SPINDLER, *User liability and strict liability in the Internet of Things and for Robots*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019, p. 139; G. WAGNER, *Robot, Inc.: Personhood for Autonomous Systems?*, in 88 *Fordham Law Review*, 2019, p. 612. L'*event data recorder* è stato reso obbligatorio, a partire dal 6 luglio 2022, per tutte le automobili di nuova immatricolazione, ai sensi del Regolamento (UE) 2019/2144 del Parlamento europeo e del Consiglio del 27 novembre 2019 relativo ai requisiti di omologazione dei veicoli a motore e dei loro rimorchi. Una misura simile potrebbe essere introdotta, per i sistemi di i.a. ad alto rischio, dal già citato *AI Act*, che all'art. 12 prevede l'installazione obbligatoria, sui suddetti sistemi, di strumenti di registrazione automatica degli incidenti.

[41] T. LEHOUILLIER E AL., *The Use of Event Data Recorders in the Analysis of Unintended Acceleration Incidents*, testo della relazione tenuta alla *23e Conférence canadienne multidisciplinaire sur la sécurité routière Montréal, Québec, 26-29 mai 2013*.

engineering (c.d. *Explainable Artificial Intelligence*)[42]. Anche in questo caso, tuttavia, dalla costante verifica empirica sarà difficile individuare una legge scientifica che consenta di spiegare – in maniera *non solo retroattiva*, ma *anche predittiva* – il rapporto eziologico tra l'*input* x e l'*output* y, proprio a causa della discrasia tra modello stocastico tipico del *machine learning* e paradigma deterministico. Se, insomma, con l'avanzamento delle conoscenze scientifiche sarà probabilmente possibile identificare, nel caso concreto, *la specifica causa dell'evento lesivo algoritmico*, più complessa sarà l'individuazione di una *legge di carattere generale* che sia in grado di spiegare il rapporto di causa-effetto tra una classe di *input* e una classe di *output*.

Questo, ovviamente, pone problemi di compatibilità con il paradigma nomologico-deduttivo dell'accertamento causale, che richiede che sia sempre individuabile una legge scientifica di copertura. La stessa sentenza Franzese[43], pur avendo segnato un primo passo nella direzione della valorizzazione del metodo induttivo, richiede pur sempre, come noto, la sussistenza di una legge scientifica di copertura, benché quest'ultima possa avere anche probabilità statistica *bassa*, qualora dal compendio probatorio emerga una spiegazione causale convincente e «la sicura non incidenza nel caso di specie di altri fattori interagenti in via alternativa» (c.d. *alta probabilità logica*).

6. L'accertamento della colpa del produttore.

La crisi del modello nomologico-deduttivo, determinata da una tendenziale inesplicabilità dei rapporti causa-effetto che governano l'agire algoritmico, sembrerebbe ripercuotersi, a cascata, sull'intera struttura del reato e, nello specifico, sull'accertamento dell'elemento soggettivo in capo al produttore[44]. In particolare, prendendo atto che l'attività illecita di quest'ultimo sarà, nella maggior parte dei casi, *involontaria*, in questa sede ci concentreremo sulla possibilità di muovergli un rimprovero colposo – nonostante non sia escluso che la condotta del produttore possa essere caratterizzata da *dolo*, e, in particolare, da dolo *eventuale*.

In assenza di conoscenze approfondite circa il concreto funzionamento degli algoritmi di *machine learning*, gli eventi lesivi scaturenti dai sistemi di i.a. possono considerarsi *prevedibili* per quanto concerne l'*an*, ma *imprevedibili* con riferimento al *quantum* e al *quomodo*.

---

[42] La letteratura in materia è orma vastissima; si vd. per tutti R. GUIDOTTI E AL., *A Survey of Methods for Explaining Black Box Models*, in 51 *ACM Computing Surveys*, 2018, p. 1 ss.; A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1110 ss; dal punto di vista normativo, vd. T. WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. Wischmeyer – T. Rademacher (eds.), *Regulating Artificial Intelligence*, Cham, 2020, p. 75 ss.; per una ricostruzione sistematica dei principali lavori in materia di *Explainable Artificial Intelligence* vd. F. SOVRANO E AL., *Metrics, Explainability and the European AI Act Proposal*, in *J*, 2022, n. 5, p. 126 ss.

[43] Cass., sez. un., 11 luglio 2022, n. 30328, Franzese.

[44] Sul problema della causalità come "origine" di tutte le altre difficoltà applicative del diritto penale nella materia del danno da prodotto vd. C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., *passim*.

L'*imprevedibilità genericamente prevedibile*[45] dei sistemi di i.a. parrebbe così mettere in discussione la stessa possibilità di muovere un rimprovero colposo al produttore, se si considera che, secondo il criterio di copertura del rischio tipico, affinché l'evento lesivo sia rimproverabile all'agente esso deve costituire realizzazione specifica del rischio che la norma cautelare mirava a prevenire. D'altra parte, in giurisprudenza, si è ormai affermato il principio in base al quale il giudizio di prevedibilità non concerne l'evento *hic et nunc* realizzatosi, quanto, piuttosto, la generica classe di eventi in cui si colloca quello oggetto del processo[46].

L'alternativa, dunque, pare netta: qualora prevalga quest'ultimo orientamento, la colpa potrebbe essere considerata sempre sussistente, dal momento che il produttore di un sistema di i.a. potrà sempre prevedere una determinata classe di eventi lesivi algoritmici; al contrario, qualora prevalga una concezione restrittiva, difficilmente potrebbe ritenersi esistente una colpa in capo al produttore, stante il carattere di intrinseca imprevedibilità dell'evento lesivo concretamente realizzatosi.

In ogni caso, un argine all'espansione del giudizio circa la prevedibilità e l'evitabilità dell'evento lesivo potrà essere costituito dal riconoscimento di un'area di rischio consentito, che possiamo definire come quell'area di impermeabilità al giudizio sulla colpa generica dell'agente, delimitata dall'esistenza di regole cautelari positive (vd. v*infra*, § 6.2.). Si pone dunque il problema di verificare: (i) se esistano cautele scritte in materia di sviluppo e commercializzazione di sistemi di i.a. (vd. *infra*, § 6.1.); (ii) quali sono i criteri che consentono l'individuazione di un'area di rischio consentito (vd. *infra*, § 6.2.).

6.1. Le regole cautelari scritte.

In assenza di generalizzazioni scientifiche sui decorsi eziologici del *decision making* algoritmico, potrebbe apparire ontologicamente impossibile la redazione di *standard* tecnici funzionali al contenimento di eventi lesivi[47]. Pur non conoscendo le precise modalità con le quali i sistemi di i.a. decidono, sappiamo tuttavia che gli errori dei sistemi di *machine learning* possono avere variamente a che fare: (i) con un errore nel codice algoritmico; (ii) con l'alimentazione dell'algoritmo con dati erronei o incompleti; (iii) con un addestramento dell'algoritmo carente o inadeguato[48]. In relazione a questi

---

[45] C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1762; vd. anche M.B. MAGRO, Robot, cyborg *e intelligenze artificiali*, in A. Cadoppi –S. Canestrari – A. Manna – M. Papa (a cura di), *Cybercrime*, Utet, 2019, p. 1208-1209; S. BECK, *Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood*, in 86 *Robotics and Autonomous Systems*, 2016, p. 139.

[46] Per una ricostruzione di tale orientamento giurisprudenziale si vd., per tutti, G. CIVELLO, (voce) *Prevedibilità e reato colposo*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, Giuffrè, 2022, p. 1017; C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio*, cit., p. 1692.

[47] Lo notava, in relazione ai prodotti tradizionali, C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 243.

[48] Va tuttavia nuovamente sottolineato che eventi lesivi potranno derivare anche da sistemi di *machine learning* pienamente conformi alle cautele esistenti, dal momento che i meccanismi probabilistici che ne

aspetti, è dunque auspicabile l'elaborazione di cautele volte a minimizzare il rischio di eventi algoritmici lesivi (c.d. *cautele improprie*)[49].

Attualmente, tuttavia, le norme cautelari scritte aventi ad oggetto lo sviluppo e la messa in commercio di sistemi di i.a. sono pochissime. Qualora venisse approvata, la già citata proposta di regolamento europeo sull'intelligenza artificiale (c.d. *AI Act*) rappresenterebbe il principale riferimento in materia. In particolare, la bozza di Regolamento europeo prevede una serie di requisiti per la commercializzazione di sistemi di i.a., specie per quelli c.d. "ad alto rischio"[50], che saranno poi specificati e concretizzati attraverso l'emanazione di norme armonizzate da parte degli enti di normalizzazione a ciò preposti (CEN, Comitato europeo di normalizzazione, e CENELEC, Comitato europeo di normalizzazione elettrotecnica)[51].

Tra i vari requisiti per lo sviluppo e la messa in commercio di dispositivi "ad alto rischio", ci limitiamo a ricordarne alcuni: l'adozione di un "sistema di gestione dei rischi" (art. 9); l'utilizzo di *dataset* "di qualità" (art. 10); la predisposizione di idonea documentazione tecnica (art. 11); l'installazione di *event data recorder* sui dispositivi (art.

---

costituiscono la base di funzionamento sono, per loro natura, *fallibili*. Sono proprio questi eventi lesivi che dovrebbero rientrare nell'area del rischio consentito.

[49] Come noto, per norme cautelari improprie si intendono quelle regole che prevedono l'adozione di misure idonee a ridurre il rischio, al contrario di quelle *proprie*, che invece tendono alla sua completa neutralizzazione. La definizione è stata proposta da P. VENEZIANI, *Regole cautelari "proprie" ed "improprie" nella prospettiva delle fattispecie colpose causalmente orientate*, Cedam, 2003, *passim*, ed è ormai comunemente accettata in dottrina, vd. per tutti C. PIERGALLINI, (voce) *Colpa (diritto penale)*, in *Enc. dir.*, Annali, X, 2017, p. 229; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, Giuffrè, 2018, p. 364; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione, La c.d. "flessibilizzazione delle categorie del reato"*, in *Criminalia*, 2012, p. 391.

[50] I sistemi di i.a. ad alto rischio sono individuati dall'art. 6 della proposta attraverso una duplice tecnica normativa: attraverso menzione espressa e attraverso rinvio al campo di applicazione di normative europee settoriali. Quanto alla prima categoria, l'art. 6, § 2 rinvia ai settori elencati all'allegato III, che comprendono, da un lato, sistemi che possono mettere in pericolo l'incolumità pubblica o la sicurezza fisica delle persone (es. n. 2 "*Gestione e funzionamento delle infrastrutture critiche*", si pensi alle componenti di sicurezza nella fornitura di acqua o di gas) e, dall'altro, sistemi che, se usati in maniera scorretta, possono causare gravi violazioni dei diritti fondamentali (es. n. 1 "*Identificazione e categorizzazione biometrica delle persone fisiche*", si pensi ai sistemi di riconoscimento facciale). Quanto alla seconda categoria, ai sensi dell'art. 6, § 1, sono altresì ad alto rischio i sistemi che soddisfano entrambe le seguenti condizioni: 1) sono prodotti – o sono destinati ad essere utilizzati come componenti di sicurezza di un prodotto – disciplinati dalla normativa di armonizzazione dell'Unione elencata nell'allegato II (si tratta, ad esempio, della Direttiva c.d. macchine CE/2006/42, del Regolamento c.d. dispositivi medici UE/2017/745, della Direttiva c.d. giocattoli CE/2009/48, della Direttiva c.d. imbarcazioni di diporto CE/1994/25); 2) sono prodotti soggetti a una valutazione di conformità da parte di terzi, ai fini dell'immissione sul mercato o della messa in servizio, ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato II. L'*AI Act* prevede una compiuta disciplina in materia di *product safety* soltanto in relazione ai suddetti sistemi "ad alto rischio" (vd. titolo III, artt. 6 ss.).

[51] La Commissione Europea, il 5 dicembre 2022, ha pubblicato la bozza di "richiesta di standardizzazione" nei confronti di CEN e CENELEC. Una volta adottata la proposta, gli *standard* dovrebbero essere pubblicati entro il 31 gennaio 2025 (dunque prima dell'entrata in vigore dell'*AI Act*, che sarà applicabile dopo 24 mesi dall'approvazione, ai sensi dell'art. 85). Per un approfondimento sul funzionamento del processo europeo di standardizzazione – che si propone di favorire una maggiore armonizzazione tra le normative interne e, di conseguenza, di agevolare la circolazione dei beni – si rinvia a Commissione europea, *La guida blu all'attuazione della normativa UE sui prodotti 2022*, 29 giugno 2022, 2022/C 247/01.

12); la previsione di meccanismi che garantiscano il controllo di una persona fisica sul funzionamento del sistema (art. 14, c.d. principio dell'*human-in-the-loop*).

L'emanazione di *standard* tecnici specifici per la messa in commercio di dispositivi intelligenti è auspicata dalla gran parte dei commentatori, come misura di attenuazione dell'incertezza regolativa in materia e, di conseguenza, del *deficit* di precisione e determinatezza della fattispecie colposa[52]. In realtà, un guadagno in termini di determinatezza sarebbe conseguibile soltanto qualora la norma positivizzata fornisse *standard* di comportamento rigidi ed esaustivi, dal momento che le cautele elastiche si fondano sul medesimo meccanismo ricostruttivo utilizzato per la colpa generica, ossia il riferimento all'agente modello[53].

La bozza di regolamento europeo sull'intelligenza artificiale, in questo senso, non pare fornire soluzioni del tutto soddisfacenti, dal momento che la vaghezza con la quale descrive i requisiti minimi per l'accesso dei sistemi di i.a. sul mercato europeo sembra prefigurare l'introduzione di norme cautelari *elastiche*. Si pensi, ad esempio, all'art. 9, § 4, lett. a), dell'*AI Act*, che stabilisce che il sistema di gestione dei rischi debba garantire l'eliminazione o la riduzione dei rischi "*per quanto possibile*": è ovvio che tale riferimento non può che aprire la strada ad una comparazione con il parametro dell'*homo eiusdem condicionis et professionis*.

Un ruolo fondamentale nella codificazione delle norme cautelari sarà probabilmente svolto, inoltre, dagli *standard* emanati dalla *International Organization for Standardization* (c.d. ISO) e dalla *International Electrotechnical Commission* (c.d. IEC). In particolare, ISO e IEC hanno istituito un *focus group* congiunto (ISO/IEC JTC 1/SC 42), finalizzato alla pubblicazione di *standard* per lo sviluppo e la produzione di sistemi di intelligenza artificiale[54]: tra 2020 e 2022, tra gli altri, sono stati pubblicati *standard* in materia di discriminazione algoritmica[55], *risk governance*[56], affidabilità[57], valutazione qualitativa dei sistemi di reti neurali artificiali[58], etc. Fondamentale sarà tuttavia la

---

[52] In generale, sul ruolo delle norme cautelari codificate nell'attenuazione del *deficit* di determinatezza della fattispecie colposa vd. per tutti G. MARINUCCI – E. DOLCINI, *Corso di diritto penale*, Giuffrè, III ed., 2001, p. 41 ss.

[53] D. CASTRONUOVO, *Responsabilità da prodotto e struttura del fatto colposo*, in *Riv. it. dir. proc. pen.*, 2005, p. 328; F. PALAZZO, *Morti da amianto e colpa penale*, in *Dir. pen. proc.*, 2011, n. 2, p. 189.

[54] Informazioni sui lavori del *focus group* – comprensive dell'elenco completo degli *standard* in materia di i.a. pubblicati fino ad oggi – sono consultabili a questo link.

[55] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24027:2021 – Bias in AI systems and AI aided decision making*.

[56] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24368:2022 – Information technology — Artificial intelligence — Overview of ethical and societal concerns*.

[57] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24028:2020 – Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*.

[58] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24029-1:2021 – Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*.

pubblicazione di *standard* che disciplinino settorialmente gli specifici campi in cui può essere utilizzata l'intelligenza artificiale.

Già *de jure condito*, d'altra parte, sono valide le norme cautelari già esistenti previste dalle direttive della c.d. *legislazione europea verticale*, a seconda del settore nel quale lo specifico dispositivo di intelligenza artificiale rientra: si pensi, ad esempio, alla c.d. Direttiva macchine (Dir. 2006/42/CE)[59], al Regolamento sui dispositivi medici (UE/2017/745)[60], alla Direttiva c.d. giocattoli (CE/2009/48)[61], etc. In chiave complementare e sussidiaria rispetto alle suddette cautele positive si pongono poi i modelli di *autonormazione* e *autocontrollo* eventualmente implementati dal produttore, che possono essere considerati dal giudice nella valutazione circa la sicurezza del prodotto (vd. art. 105, co. 3, d.lgs. 6 settembre 2005, n. 206, cod. cons., che fa riferimento, tra i parametri di valutazione della sicurezza del prodotto, anche ai «codici di buona condotta in materia di sicurezza vigenti nel settore interessato»).

6.2. Il rapporto tra regole cautelari scritte e regole cautelari non scritte: quale spazio per il rischio consentito?

Poniamo ora il caso che il dispositivo intelligente abbia cagionato un evento lesivo *nonostante l'osservanza*, da parte del produttore, delle norme cautelari codificate. Il rispetto delle cautele positive stabilite dagli *standard* tecnici non potrà infatti evitare il rischio che dal concreto operare dei sistemi di i.a. derivino eventi lesivi per la vita, la salute o l'integrità fisica: nel momento in cui, infatti, si attribuisce una capacità di autonomo *decision making* ad un algoritmo – ancorché parziale e sotto supervisione di un essere umano – non si potrà escludere che da questo si verifichino decorsi eziologici dannosi. Di qui la questione: il rispetto degli *standard* tecnici di settore – quando saranno introdotti – sarà sufficiente ad escludere la colpa in capo al produttore, o il giudice dovrà invece valutare se la condotta di questi sia conforme a quella dell'*homo eiusdem condicionis et professionis*?

È un problema ben noto, che ancora non ha trovato una soddisfacente e condivisa soluzione sul piano dogmatico. Tradizionalmente, come noto, si ritiene che il rispetto delle norme cautelari positive non esoneri l'agente dal tenere un comportamento diligente, da identificarsi sulla base del criterio del c.d. agente modello[62]. Nel caso delle attività produttive, tuttavia, il problema si complica, poiché le norme cautelari, oltre che

---

[59] Direttiva 2006/42/CE del Parlamento europeo e del Consiglio, del 17 maggio 2006, relativa alle macchine. La direttiva è attualmente oggetto di una proposta di riforma presentata dalla Commissione Europea (Proposta di Regolamento del Parlamento europeo e del Consiglio sui prodotti macchina; COM (2021) 202 final, 21 aprile 2021).

[60] Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici.

[61] Direttiva 2009/48/CE del Parlamento europeo e del Consiglio del 18 giugno 2009 sulla sicurezza dei giocattoli.

[62] Vd. per tutti G. MARINUCCI, *La colpa per inosservanza di leggi*, Giuffrè, 1965, p. 236; F. MANTOVANI, (voce) *Colpa*, in *Dig. pen.*, II, Utet, 1988, p. 309.

una funzione di *prevenzione* rispetto al verificarsi dell'evento lesivo, hanno altresì una funzione di *garanzia* e di *orientamento* nei confronti dell'agente, in settori particolarmente rischiosi e ad alta complessità tecnico-normativa.

Si apre così la questione circa la sussistenza di un'area di c.d. "rischio consentito" (*erlaubtes risiko*), da tempo dibattuta in dottrina[63] e oggetto di rinnovata attenzione, recentemente, proprio in relazione agli eventi lesivi derivanti da sistemi di i.a.[64].

Non è questa la sede per soffermarci sulle varie ricostruzioni ermeneutiche del "rischio consentito" che, nel corso del tempo, sono state date. Possiamo qui limitarci a citarne una particolarmente autorevole, proposta da Gabrio Forti, secondo la quale la nozione di rischio consentito indicherebbe un'area di «condotte pericolose, ammesse dall'ordinamento nonostante che l'adozione di cautele idonee a contrastare i possibili svolgimenti lesivi sia destinata a residuare un certo grado di pericolosità»[65]. In quest'accezione, il rischio consentito viene anche qualificato come "rischio residuale *(Restrisiko)*"[66], intendendosi, con tale designazione icastica, quel *pericolo marginale* che le misure preventive non sono in grado di disinnescare (o che, sulla base di un rapporto costi-benefici, le autorità che "producono" la normativa cautelare *decidono* di non disinnescare) e che, di conseguenza, ricade sulla società. In quest'accezione, la nozione di "rischio consentito" delinea quelle ipotesi in cui il rispetto delle norme cautelari *codificate* impedisce che possa essere mosso all'imputato un rimprovero per non aver osservato norme di diligenza, prudenza, perizia *non codificate*.

---

[63] La letteratura sul tema è ormai di ampiezza considerevole; si vd. G. FORTI, *Colpa ed evento nel diritto penale*, Giuffrè, 1990, p. 250 ss.; ID., *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, in *Criminalia*, 2006, p. 155 ss.; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., *passim;* ID., *Il paradigma della colpa nell'età del rischio*, cit., p. 1670 ss.; V. MILITELLO, *Rischio e responsabilità penale*, Giuffrè, 1988, p. 55 ss. (che preferisce, tuttavia, l'espressione "rischio adeguato"); F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006, p. 227 ss.; F. CONSULICH, (voce) *Rischio consentito*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1102 ss.; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 335 ss.; D. CASTRONUOVO, *Principio di precauzione e beni legati alla sicurezza*, in *Dir. pen. cont.*, 21 luglio 2011, p. 1 ss.; ID., *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, Aracne, 2012; M. DONINI, *Il volto attuale dell'illecito penale*, Giuffrè, 2004, p. 119 ss.; A. MASSARO, *Principio di precauzione e diritto penale:* nihil novi sub sole?, in *Dir. pen. cont.*, 2011; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, in *Dir. pen. e proc.*, 2008, p. 647; C. RUGA RIVA, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, in E. Dolcini – C.E. Paliero (a cura di) *Studi in onore di Giorgio Marinucci*, II, Giuffrè, 2006, p. 1743 ss.

[64] Vd., con vari accenti, S. GLESS – E. SILVERMAN – T. WEIGEND, *If Robots Cause Harm, Who Is to Blame: Self-Driving Cars and Criminal Liability*, in *New Criminal Law Review*, 2016, pp. 430-431; C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1750; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 2021, n. 1, p. 116 ss.; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (a cura di), *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, pubblicato anche in *disCrimen*, 15 maggio 2020, p. 5; A. CAPPELLINI, Machina delinquere non potest?, cit., p. 19; A. FIORELLA, *Responsabilità penale dei* Tutor *e dominabilità dell'Intelligenza Artificiale, Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 656 ss.; D. PIVA, Machina discere, (deinde) delinquere et puniri potest, *ibidem*, p. 681 ss.

[65] G. FORTI, *Colpa ed evento nel diritto penale*, cit., p. 457.

[66] G. FORTI, *Colpa ed evento nel diritto penale*, cit., p. 457.

Come noto, l'istituto del rischio consentito è strutturalmente legato all'idea che la norma cautelare è *inidonea a prevenire tutti i rischi che è finalizzata a prevenire*. Si tratta di un'inidoneità *prevedibile* e *programmata*: il legislatore – e gli altri soggetti che, a vario titolo, sono deputati alla definizione delle regole cautelari – nel momento in cui delineano il bilanciamento tra interessi divergenti *sanno* che la cautela è talvolta destinata a fallire, poiché non sarà in grado di avere efficacia impeditiva nei confronti di *tutti* gli eventi lesivi[67]. Così, per riportare il discorso sui binari dell'intelligenza artificiale, il legislatore europeo, con l'*AI Act*, introduce una serie di norme cautelari volte a *minimizzare* il rischio di eventi lesivi algoritmici, non a *neutralizzarlo*: accetta, dunque, sebbene implicitamente, che dallo sviluppo e dalla commercializzazione di sistemi di i.a. derivi una *certa quota* di eventi offensivi.

L'unico margine di operatività di una valutazione della colpa generica dell'agente dovrebbe configurarsi nel caso in cui la regola cautelari manifesti *segnali univoci* di fallimento "non preventivato"[68]. In questi casi – quando la norma cautelare scritta, dunque, si riveli inidonea al raggiungimento degli obiettivi che ne costituiscono il fondamento –, la diligenza impone di adottare cautele ulteriori; il che, nella prospettiva del produttore di un sistema di i.a., potrebbe significare: il fornire ulteriori informazioni al consumatore, l'aggiornamento del *software*, o, nei casi più gravi, il ritiro o il richiamo del prodotto.

Il problema che si è sempre posto nell'applicazione di questi principi, tuttavia, è quello di riconoscere univocamente i suddetti *segnali* di fallimento "non preventivato" delle regole cautelari scritte. Nel settore dell'intelligenza artificiale, un aiuto in questo senso potrebbe derivare dalla possibilità, per i produttori, di monitorare costantemente le *performance* degli algoritmi, attraverso i già citati meccanismi di *explainability* – possibilità che potrebbe determinare un incremento nella *calcolabilità* degli errori e, di conseguenza, degli eventi lesivi che possono derivarne.

Poniamo il caso, ad esempio, che il legislatore stabilisca un apparato di norme cautelari rigide per lo sviluppo e la produzione di *software* di guida autonoma (es. modelli di raccolta e gestione dati, addestramento, sperimentazione, etc.), esplicitando che la fissazione di tali norme è funzionale a mantenere il tasso di incidenti al di sotto della soglia dell'x% rispetto all'utilizzazione del veicolo. Nel caso in cui, nonostante il rispetto delle norme cautelari rigide, il produttore verifichi che il tasso di sinistri derivanti dall'utilizzo del *software* sia superiore rispetto alla soglia del rischio lecito, questi dovrà adottare cautele ulteriori per minimizzare i rischi e, eventualmente, richiamare o ritirare il prodotto. Non è necessario, a tal fine, che la suddetta verifica abbia i crismi di un rigoroso accertamento del nesso di causalità condotto con gli *standard* penalistici – accertamento che, come abbiamo rilevato, sconta problematiche che potrebbero risultare insormontabili. La verifica dovrebbe essere tuttalpiù in grado di *isolare* gli errori algoritmici, senza che sia invece necessario, in ipotesi, individuare il

---

[67] F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1116.

[68] Parlano di eventuale "fallimento" delle norme cautelari, tra gli altri, G. FORTI, *Colpa ed evento nel diritto penale*, cit., p. 671 ss.; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 61 ss., S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380.

preciso *input* che ha determinato il verificarsi dell'errore algoritmico e, dunque, dell'evento lesivo.

Il superamento del tasso stabilito in via legislativa non determinerebbe l'automatica imputazione degli eventi lesivi "ulteriori" al produttore, ma dovrebbe essere considerato come segnale di *incipiente fallimento delle norme cautelari scritte*: dovrebbe, dunque, indurre il produttore ad adottare cautele più pregnanti rispetto a quelle previste dalla normativa esistente.

Un approccio simile è stato proposto da una parte della dottrina civilistica, con riferimento al tema contiguo dell'accertamento di un difetto di *design*[69]. In particolare, tale orientamento ipotizza di valutare la difettosità del prodotto intelligente attraverso la comparazione tra la prestazione complessiva del sistema di i.a. che ha causato il danno e la *performance* di un modello di algoritmo mediamente sicuro. A tal fine, si prospetta l'opportunità di individuare una soglia di riferimento – relativa al rapporto percentuale tra eventi lesivi verificatisi e utilizzo complessivo del *software* – al di sopra della quale l'algoritmo dovrebbe essere considerato difettoso.

Simili soluzioni – sia nel settore civile, sia in quello penale – avrebbero il vantaggio di spostare il *focus* della valutazione sulla sicurezza del prodotto intelligente, dalla prestazione singola alla sua *performance* complessiva: una traslazione valutativa che appare indispensabile se si vuole godere dei benefici derivanti dallo sviluppo dei sistemi di i.a., senza che singoli eventi lesivi possano mettere in discussione la sicurezza del prodotto.

In conclusione, non possono, in ogni caso, essere trascurati i limiti di una siffatta prospettazione. Potrebbe emergere, innanzitutto, un limite tecnico, relativo alla disponibilità, per il produttore, di affidabili e tempestivi meccanismi di controllo sulle *performance* algoritmiche: tale problema, tuttavia, potrebbe essere risolto in un futuro non troppo lontano, dati i continui progressi scientifici in materia di *Explainable AI*. Più pregnante è invece la questione di natura politica: quale regolatore, infatti, espliciterebbe mai in maniera così diretta il bilanciamento tra esercizio di un'attività pericolosa e beni giuridici di estremo rilievo (quali, in ipotesi, vita ed integrità fisica)? Il rischio, d'altra parte, è che anche qualora si pervenisse all'individuazione di una soglia quantitativa, questa sia ispirata a logiche *iper-cautelative*, che determinerebbero di fatto, per il produttore, la necessità di adottare sempre cautele *ulteriori* rispetto a quelle positive.

7. Una tutela anticipata in relazione ai sistemi di i.a. pericolosi: prospettive *de jure condito* e *de jure condendo*.

Nonostante lo sforzo di ricostruzione ermeneutica sin qui svolto, il cammino per il riconoscimento di una responsabilità penale in capo al produttore di sistemi di i.a. appare in tutta la sua tortuosità. In particolare, l'imprevedibilità e l'opacità del *decision*

[69] J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 63 ss.

making algoritmico *allontanano* l'evento lesivo dalla condotta umana, incidendo così sulla possibilità di muovere un rimprovero colposo al produttore e, prima ancora, di individuare con certezza il nesso di causalità tra condotta umana ed evento lesivo. Già da tempo, d'altra parte, autorevole dottrina segnala come il diritto penale, di fronte alle sfide della società del rischio tecnologico, si trovi di fronte ad una vera e propria "crisi da incontenibilità"[70], determinata dalla difficoltà di ritenere nel "tipo" fenomeni complessi, globalizzati, imperscrutabili.

Di qui la convinzione che – a meno di non voler rinunciare ai principi cardine del diritto penale – una responsabilità del produttore per gli eventi lesivi algoritmici costituirà una rarità. In questo contesto, una tutela penalistica – la sola, ci pare, in grado di esprimere il disvalore insito nella creazione o nel mantenimento di rischi illeciti a beni giuridici di primaria importanza, quali la vita e l'integrità fisica – potrebbe appuntarsi in forma anticipata, attraverso il ricorso a reati colposi di mera condotta.

Già *de jure condito* si potrebbe fare riferimento alle fattispecie contravvenzionali previste dall'art. 112 cod. cons., commi 1, 2, 3, che puniscono la commercializzazione di prodotti *pericolosi*, e che trovano applicazione in via sussidiaria, «*se il fatto non costituisce più grave reato*»: l'efficacia deterrente di tali norme, tuttavia, rischia di essere risibile, stante l'esiguità delle pene ivi previste (la pena per la fattispecie più grave, quella di cui al primo comma, prevede l'arresto da sei mesi ad un anno e l'ammenda da 10.000 a 50.000 euro). Non è detto, d'altra parte, che il sistema di i.a. rientri sempre nella nozione di "prodotto" fornita dall'art. 3, co. 1, lett. e), cod. cons., in base alla quale per "prodotto" si intende «qualsiasi prodotto destinato al consumatore, anche nel quadro di una prestazione di servizi, o suscettibile, in condizioni ragionevolmente prevedibili, di essere utilizzato dal consumatore, anche se non a lui destinato, fornito o reso disponibile a titolo oneroso o gratuito nell'ambito di un'attività commerciale […]». Tralasciando l'evidente problema di circolarità della norma (*un prodotto è… un prodotto!*), la riconducibilità dei sistemi di i.a. alla suddetta definizione sembrerebbe possibile soltanto nei casi in cui l'algoritmo di intelligenza artificiale sia incorporato in un bene *destinato al consumatore* o *suscettibile di essere da lui utilizzato* (es. sistema di guida autonoma installato su un veicolo)[71]. Non sembrerebbe, invece, esservi alcuno spazio per un'estensione dell'ambito applicativo della norma ai c.d. *prodotti digitali*, ovverosia ai *software* che non sono incorporati in un *hardware*.

Uno spunto per l'introduzione di una tutela penale anticipata potrebbe venire, piuttosto, dalla già citata Proposta di Regolamento europeo, che – nel prevedere una serie di requisiti per lo sviluppo e la messa in commercio di sistemi di i.a. – obbliga gli Stati membri ad introdurre sanzioni «effettive, proporzionate e dissuasive» per il caso di mancato rispetto della disciplina ivi stabilita (vd. art. 71, § 1, *AI Act*). Una parte della dottrina, a tal proposito, ha proposto l'introduzione di reati colposi di mera condotta, volti a criminalizzare: (*i*) l'omessa predisposizione, da parte del produttore, di

---

[70] C.E. PALIERO, *L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici*, in *Riv. it. dir. proc. pen.*, 1994, cit., p. 1238.

[71] F. CONSULICH, Flash offenders. *Le prospettive di* accountability *penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.*, 2022, fasc. 3, p. 1039.

determinati presidi di sicurezza nei sistemi di i.a.; (*ii*) la disattivazione, la mancata attivazione o il mancato aggiornamento, da parte dell'utilizzatore, dei presidi di sicurezza di cui invece il sistema di i.a. fosse stato originariamente dotato[72].

In quest'ottica, uno spazio potrebbe essere riservato, sempre *de jure condendo*, alla valorizzazione del modello ingiunzionale[73], favorendo così un coordinamento tra sanzione penale e capillare controllo amministrativo[74]. L'ingiunzione, da parte delle autorità competenti, di assumere misure di sicurezza (es. adozione di ulteriori cautele, sottoposizione dei dispositivi intelligenti a *training* o sperimentazione aggiuntiva, fino all'ordine di disattivazione) potrebbe infatti essere presidiata da sanzioni penali, così da fronteggiare – in un'ottica di governo condiviso del rischio – l'eventuale inosservanza, da parte dell'impresa, di prestazioni di *facere* individualizzate e infungibili[75]. Non si può non rilevare, tuttavia, come misure di questo tipo – già presenti nella disciplina generale sulla sicurezza dei prodotti (vd. art. 112, co. 1 e 3, cod. cons.) – siano, ad oggi, rimaste soltanto sulla carta.

Infine, in una logica di "democratizzazione" dei processi di valutazione e prevenzione del rischio[76] – e prendendo atto delle conoscenze specifiche esistenti all'interno delle imprese –, si potrebbero prevedere, a carico delle società produttrici, degli *obblighi di diffusione delle informazioni* sui rischi derivanti dallo sviluppo e dalla messa in commercio dei sistemi di i.a.[77].

8. Conclusioni.

L'ampio e stimolante dibattito emerso durante questo Corso testimonia come l'intelligenza artificiale ponga davanti ai penalisti – e, in generale, a tutti i giuristi – problemi inediti e urgenti. D'altra parte, se è vero che ci troviamo di fronte ad un fenomeno dalle caratteristiche nuove e dirompenti, lo studio dell'impatto dell'intelligenza artificiale sulla responsabilità penale sembrerebbe costituire un avamposto privilegiato per osservare alcune tendenze più generali del diritto penale,

---

[72] Così F. CONSULICH, Flash offenders, cit., p. 1038-1039; vd. anche F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective*, cit., p. 459.

[73] C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1773.

[74] In generale, sulla tecnica ingiunzionale vd. G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.*, 2005, fasc. 1, p. 56 ss.; A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, Monduzzi, II ed., 2000, p. 50 ss.

[75] È questo uno dei più apprezzabili vantaggi del modello ingiunzionale; vd. a tal proposito A. ALESSANDRI, *Parte generale*, cit., p. 51-52.

[76] G. FORTI, *"Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, cit., p. 217.

[77] È la tesi di G. FORTI, *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, cit., *passim*; per un "aggiornamento" della tesi al contesto produttivo dell'intelligenza artificiale vd. C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1773.

relative, in particolare, all'accertamento del nesso di causalità e della colpa in contesti di incertezza scientifica.

La riscontrata difficoltà nell'individuazione delle responsabilità *personali* nell'ambito della produzione dei sistemi di i.a. sembrerebbe inoltre aprire ad una rinnovata riflessione sul ruolo della responsabilità amministrativa degli enti nella prevenzione delle attività criminali[78]. In particolare, nel settore della produzione dei sistemi di i.a. emerge plasticamente la disarmonia tra carattere intrinsecamente *plurisoggettivo* degli illeciti e conformazione *personalistica* della responsabilità penale: una disarmonia che si riscontra in relazione a tutti i fenomeni tipici della criminalità d'impresa e che sconta il rischio che la ricerca del soggetto rimproverabile all'interno delle società si trasformi in una caccia al capro espiatorio[79]. *De jure condendo*, proprio il settore della produzione di sistemi di i.a. – per le peculiarità che abbiamo tentato di descrivere – potrebbe rappresentare un utile banco di prova per sperimentare modelli volti a realizzare una maggiore *autonomia* della responsabilità dell'ente rispetto a quella della persona fisica[80].

Per concludere, resta sullo sfondo l'interrogativo su quale ruolo il diritto penale potrà concretamente svolgere, sul lungo periodo, in una società retta da meccanismi artificiali, il cui funzionamento non è del tutto conoscibile nemmeno ai loro creatori. La sfida sarà quella di salvaguardare l'insostituibile *utilità sociale* del diritto penale, senza rinunciare alle garanzie e ai principi che ne costituiscono il fondamento democratico (colpevolezza e legalità, *in primis*).

---

[78] B. PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, cit., p. 362 ss.

[79] Su tale rischio vd. G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 57. Per una recente analisi, da una prospettiva sociologica, del meccanismo del capro espiatorio organizzativo vd. M. CATINO, *Trovare il colpevole. La costruzione del capro espiatorio nelle organizzazioni*, il Mulino, 2022.

[80] Sull'emancipazione, *de jure condendo*, della responsabilità *ex* d.lgs. 231/2001 dalla dipendenza dal reato della persona fisica vd. A. GARGANI, *Profili della responsabilità collettiva da reato colposo*, in *Riv. trim. dir. pen. econ*. 1-2/2022, p. 48 ss., e, in tema di reati ambientali, L. MALDONATO, *Il crimine ambientale come crimine delle corporations: cooperazione pubblico-privato e responsabilità indipendente dell'ente*, in *Riv. trim. dir. pen. econ*., 3-4/2021, p. 504 ss.; per una valorizzazione, *de jure condito*, dell'art. 8 del d.lgs 231/2001, vd., sempre in materia di criminalità ambientale, F. CONSULICH, *Il giudice e il mosaico. La tutela dell'ambiente, tra diritto dell'Unione e pena nazionale*, in *Leg. pen*., 27 luglio 2018, p. 22 ss.; in materia di imputazione di eventi lesivi algoritmici, ID., *Il nastro di Möbius. intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, cit., p. 224 ss.