QUALE IMPATTO DELL'INTELLIGENZA ARTIFICIALE SUI PROFILI DI RESPONSABILITÀ PENALE?

Errare humanum fuit

1. PREMESSE E PRIME DEFINIZIONI

Il progresso tecnico della robotica segna l'entrata nell'era dell'intelligenza artificiale, la scoperta del fuoco del XXI secolo, prettamente caratterizzata dall'evoluzione fulminea, e, soprattutto, dal senso di disorientamento che ha travolto (e che ancora travolge) tutti coloro che attivamente o passivamente vi entrano in contatto. Popperianamente, si sta assistendo alla distruzione del tempio del sapere, che, fino ad ora edificato su certezze cristallizzate ed immobilizzate, sta adesso vivendo una ri-costruzione per il tramite di un pensiero scientifico "vivo", ossia che si corregge e si trasforma di continuo.

Rosenberg avrebbe forse parlato di "scatola nera", criticando la statica della innovazione[1], di cui nulla si conosce e per la quale si provano sentimenti contrastanti che oscillano fra il brio del nuovo e dell'utile alla vita di ciascuno ed il senso di agitazione e di inadeguatezza che ne derivano, fino all'espressione di una rivoluzione epidemica e a quella di una prossima frattura antropologica.

Privi (ed anche privati) di strumenti definitori adeguati alla comprensione del nodo problematico - e pur tralasciandosi una, pur potenzialmente interessante disamina sulla concezione della *intellegentia*[2]- diremo che rientra nella intelligenza artificiale la capacità (dei sistemi tecnologici) di agire adattandosi ad una serie indefinita ed infinita di elementi circostanziali, assumendosi

[1] N. Rosenberg, *Inside the black box: technology and economics*, 2004. V. p. 141

[2] B. Fragrasso, *Intelligenza artificiale e responsabilità penale. Principi e categorie alla prova di una tecnologia 'imprevedibile',* Torino, 2025. V. p. 19. Come può parlarsi di intelligenza artificiale, stante le innumerevoli concezioni di *intelletibile* a cui poter aderire? Intelligenza è la capacità di emulare un comportamento umano o di agire razionalmente (*human vs rational intelligence*)? Intelligenza è da correlarsi ad un prodotto del pensiero o basterebbe la esternazione di comportamento nell'ambiente esterno?

decisioni (sottoforma di *output* impattanti) e massimizzandosi esponenzialmente il risultato atteso[3].

Il passo successivo è quello di circoscrivere i modelli che rientrano in una siffatta definizione per tenersi a mente che «*as soon as it works, no one calls it AI anymore*[4]».

Quel che ancor qualche anno fa erano contegni[5] umani (da intendersi in lato senso: non solo gesti meccanici, ma espressioni e condotte con sfumature più profonde, tipicamente umanizzate ed umanizzanti), oggi "*si tengono*" con paradigmi connessionisti e probabilistici per il tramite delle macchine artificiali[6].

Trattando dei c.d. *machine learning*, il cui tratto distintivo è l'essere *unpredictable by design*[7] (imprevedibili e spiazzanti nelle esternalizzazioni derivate[8]) quello che interessa inserire nel vetrino portaoggetti da vagliare ai fini ultimi della comprensione del problema penale pare così presentare tre caratteristiche specifiche: autonomia, automazione ed opacità[9].

La conseguenza è il senso di manchevolezza e di inidoneità a tratti da cui resta colto il giurista, forzato ad un bilanciamento fra il mondo fuori dalle aule di giustizia, quello del progresso tecnico e scientifico in piena evoluzione da assecondarsi e controllare (da lontano) ed i fatti interni a quelle stanze dove ancora vige il principio della ponderazione, dello studio lento ed attento degli atti.

---

[3] *Ibidem.* Questo è il risultato dato dalla adesione al c.d. *standard model*, cioè l'agire razionalmente conseguendo obiettivi determinati e specifici pur se in ambienti complessi ed incerti.

[4] *Ivi.* Cit. p. 27

[5] Contegno /*kon'teɲo*/ s. m. [der. del lat. *continere* "*contenere*"]. - 1. [*modo con cui una persona sta o si atteggia*] ≈ atteggiamento, comportamento, condotta, maniere, modo di fare. 2. [*assol., comportamento serio o serioso*] ≈ compostezza, decoro, riserbo, ritegno, serietà. Così nella Treccani < https://www.treccani.it/vocabolario/contegno_(Sinonimi-e-Contrari)/ >

[6] A che prezzo poi, quello è un nodo che non riguarda questo contributo. Certo è, che per quanto ci si provi rendere un algoritmo '*person in Recht*', il prodotto finale è una depersonalizzazione del lavoratore e con risultati di dubbia qualità. In un provvedimento del tribunale di Milano, di condanna di un avvocato che ha usato i sistemi di IA per redigere contributi, si legge: «[della] *scarsa qualità degli scritti difensivi e dalla totale mancanza di pertinenza o rilevanza degli argomenti utilizzati; l'atto è infatti composto da un coacervo di citazioni normative e giurisprudenziali astratte, prive di ordine logico ed in gran parte inconferenti rispetto al thema decidendum ed, in ogni caso, tutte manifestamente infondate*».

[7] *Ibidem.* V. pp. 332 e ss. Non è una connotazione patologica-funzionale della tecnologia, ma è quello che i disegnatori si auspicavano di ottenere in termini di prestazioni massime ed efficienza creativa. Sulla c.d. *black box*, v. p. 56 e ss.

[8] *Ibidem,* v. p. 33. Insomma, la impossibilità di poter applicare il broccardo latino "*cum hoc, ergo propter hoc"*.

[9] B. Fragrasso, *Intelligenza artificiale.* Op. cit. v. Pp. 47 e ss. Si considerano ai fini della digressione sui profili della responsabilità penale, solo i meccanismi di cui sopra, in quanto autonomi, nel senso più immediato del termine, e quindi non suscettibili di alcuna forma di controllo dal lato umano sui processi di *output*.

La domanda "C*an machines learn*?" ha subìto una dilatazione di significante ed oggi si ripresenta come: "*Can machines do wrongs*?".

Questa la prospettiva da cui muovere per il penalista. L'analisi dell'impatto che le nuove forme di agire dis-umano o meta-umano potrebbero avere in termini di imputazione della responsabilità penale; premettendosi che l'agire dis-umano si fonda sul «*sul mancato dominio di un fatto offensivo effettivamente dominabile*[10]».

Il dibattito accademico si interroga così su chi – e a quali condizioni – debba rispondere penalmente degli errori o danni causati da sistemi di IA avanzati.

Fin dalle prime riflessioni, è emerso che l'ordinamento penale è intrinsecamente antropocentrico: quale autore del reato presuppone un essere umano dotato di volontà, coscienza e capacità di colpevolezza[11]. Il che porta ad una esclusione della concepibilità in capo alla macchina di una soggettività penale autonoma, tanto da potersi dire che *machina delinquere et puniri non potest.*

Si svilisce perciò il senso dell'intelligenza artificiale, che, per quanto irrompente nella quotidianità, è non solamente inferiore nel sentire, percepire, desiderare e volere, ma apparentemente *instrumentum* fra i pollici opponibili degli umani. L'uomo parrebbe fautore di tutto quello che lo circonda, eppure, contraddizione fattuale che resta quella in una vita concepita come *Onlife*[12]*,* proprio l'essere umano è in balìa delle emozioni, degli altri uomini e delle macchine. '*Condivide con*' questi le azioni, gli illeciti, ma non la responsabilità penale. Sarebbe infatti impensabile riconoscere una qualsiasi forma di responsabilità per l'evento lesivo accorso ad un "contegno" – esteriorizzato artificialmente- di un dispositivo insensibile e non senziente[13].

Un ostacolo insormontabile quello del principio di colpevolezza, di personalità, della presunzione di innocenza: '*l'uomo è misura di tutte le cose',* anche di quelle non sensienti, ed è risposta attualmente del problema che tange l'individuazione del responsabile per gli esiti dannosi che sono prodotti dalla tecnologia[14].

Fatti dannosi e lesivi che nel reale potrebbero avere come fautore unico proprio un sistema tecnologico, porrebbero il legislatore di fronte ad un vuoto di tutela che è noto con l'inglesismo:

---

[10] A. Fiorella, *Responsabilità penale* (voce), in Enc. Dir., vol XXXIX, 1988, p. 1289. In questo caso c'è la inadeguatezza non solo perché il dominio sui fatti non c'è, ma anche perché 'l'autore' del reato è una macchina.

[11] F. Basile, B. Fragasso, *Intelligenza artificiale e diritto penale: prove tecniche di convivenza*, in Dirittodidifesa.eu, 2023 – sul principio "*machina delinquere non potest*" e antropocentrismo del diritto penale

[12] F. Basile, *Diritto Penale e intelligenza artificiale. "Nuovi scenari",* Giuliano Balbi - Federica De Simone Andreana Esposito - Stefano Manacorda (a cura di), 2022. V. p. 12

[13] *Res* terza, fra soggetto (che agisce) ed oggetto (che ha agito). *Position Paper – Responsabilità penale e intelligenza artificiale*, Associazione Italiana Giovani Penalisti, luglio 2025, p. 4 – sull'assenza di soggettività penale autonoma delle IA e la ricostruzione della responsabilità umana mediata.

[14] Costituzione della Repubblica Italiana, art. 27, commi 1-2-3

responsibility gap[15], strettamente correlato all'opaco funzionamento e modo di essere intrinseco dell'algoritmo, paralizzando il giudizio di imputazione sul piano penale.

In tale prospettiva introduttiva pare utile delineare (anzitutto) il quadro normativo di riferimento – sia sovranazionale che nazionale – relativo all'uso dell'IA, per poi concentrare l'analisi sui profili penalistici: separatamente, la posizione dell'utilizzatore di sistemi di IA e quella del produttore/sviluppatore, considerandosi sia le ipotesi di condotte dolose che quelle colpose.

1.1 Principi Generali - Il punto di partenza

La Costituzione italiana, all'art. 27 commi 1-3, sancisce noti principi cardine in materia penale. *In primis*, «*la responsabilità penale è personale*». Il che implica il divieto di punire un soggetto per un fatto altrui e postula il principio di colpevolezza: ogni reato richiede una condotta umana colpevole (dolo o colpa), escludendo forme di responsabilità oggettiva[16].

Vige inoltre la presunzione di innocenza fino a condanna definitiva (art. 27 co. 2 Cost.).

Sul piano sanzionatorio, la pena non può consistere in trattamenti contrari al senso di umanità e deve tendere alla rieducazione del condannato (art. 27 co. 3 Cost.): la funzione rieducativa-reintegrativa è fine primario della pena, accanto alla prevenzione e alla retribuzione.

Questi principi generali pongono sfide nuove e peculiari quando si tratta di Intelligenza Artificiale (IA).

Si premetta che la soggettività di tali meccanismi è da escludersi, per una duplice motivazione: non vi può essere responsabilità se non vi è autodeterminazione, né presa di consapevolezza del valore semantico e convenzionale dei propri contegni; non c'è pena per chi non può essere chiamato a rispondere (nei termini più lati) di stigmatizzazione e sofferenza, venendo di fatto annientato quel senso di rieducazione e socializzazione, che poi è fine ultimo e nobile della pena[17].

In sintesi, un sistema di IA, per quanto autonomo, non possiede coscienza, volontà colpevole, né potrebbe mai essere destinatario di una pena rieducativa in senso proprio[18]. Pertanto, qualsiasi illecito derivante dall'uso di IA va ricondotto alla persona umana che ha progettato, utilizzato o non adeguatamente controllato il sistema e che *in extremis* potrebbe essere rimproverata in segno

[15] 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
[16] G. Leone, *Il reato aberrante,* Napoli, 1940. A riguardo, v. P. 139: «[l'evento] *non ha alcuna relazione soggettiva coll'agente oppure ha col medesimo quella modesta, elementare relazione soggettiva che, mentre non basta a dar luogo alla responsabilità per colpa, serve di base, secondo l'orientamento dei negatori della responsabilità senza colpa, ai casi di c.d. responsabilità oggettiva*».
[17] B. Fragrasso, *Intelligenza artificiale.* Op. cit. v. p. 3
[18] *Ibidem*

di biasimo per quanto fatto o non fatto[19]. Ed è la stessa previsione dell'art. 27 Cost. ad imporre di mantenere come polo attrattivo ai fini dell'imputazione penale il fattore umano, evitandosi scorciatoie che configurino responsabilità penali oggettive o attribuzioni punitive a entità non umane.

Breve: il diritto penale è edificato sull'essere umano (personologico e personocentrico), sul rimprovero personale e colpevole, sul giudizio scaturente dall'uomo che brancola nel 'crepuscolo del dubbio'.

2. QUADRO NORMATIVO: LE FONTI EUROPEE ED IL DIRITTO NAZIONALE

Normativa UE (il c.d. *AI Act*)

A livello sovranazionale, l'Unione Europea ha adottato il Regolamento (UE) 2024/1689 (c.d. AI Act), entrato in vigore il 1º agosto 2024[20], e che rappresenta il primo quadro giuridico vincolante[21] nel sistema europeo che abbia ad oggetto sistemi artificiali e ad esser improntato al *risk-based approach.*

Si classificano i sistemi di IA in categorie di rischio (inaccettabile, alto, limitato, minimo) e si vietano taluni usi considerati incompatibili con i diritti fondamentali[22].

Nello specifico, sono proibiti sistemi di *social scoring* atti a manipolare il comportamento umano sfruttandone vulnerabilità e l'uso di IA per il c.d. *predictive policing,* basato unicamente su profilazione o caratteristiche personali. È in linea di massima bandito anche l'impiego del riconoscimento biometrico in tempo reale da parte delle forze dell'ordine, salvo eccezioni strettissime[23].

---

[19] A. Cappellini, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale,* in Criminalia, 27.03.2019 < https://discrimen.it/wp-content/uploads/Cappellini-Machina-delinquere-non-potest.pdf>

[20] v. *Intelligenza Artificiale: approvata la legge delega italiana (ddl AI)* <https://www.diritto.it/approvata-legge-delega-italiana-intelligenza-artifi/ >

[21] A. Bradford parla di *Brussels effect*; in termini esplicativi, la *pole position* europea, che in fatto di supremazia tecnologia, parrebbe porsi in termini di ossimoro, se messa a confronto con USA e Cina, ha come sostrato l'idea che a livello globale, le imprese di tutto il mondo, non solo non possono rinunciare al mercato dei consumi europeo, ma, devono per questo motivo sottostare alle regole e agli standard dalla stessa Europa posti in essere. Se il *de facto* non bastasse, la linea guida europea è stata tracciata con demarcazione anche *de iure*, si veda l'art. 2, par. 1, lett. C) che disciplina l'applicazione del regolamento di cui sopra, agli utilizzatori profani e professionalizzati ovunque situati.

[22] *Artificial Intelligence Act: MEPs adopt landmark law | News | European Parliament* < https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law >

[23] *Ibidem*

Per i sistemi ad alto rischio (es. applicazioni tecnologiche in settori critici come sanità, trasporti autonomi, istruzione, occupazione, forze di polizia, giustizia), l'*AI Act* impone rigorosi obblighi a carico dei fornitori e dei distributori: dalla valutazione e mitigazione dei rischi, all'alta qualità dei dati di addestramento, registrazione e tracciabilità degli eventi (*logging*), trasparenza verso gli utenti e soprattutto garanzia di controllo umano (*human oversight*) sul funzionamento del sistema[24].

Un aspetto – l'obbligo di predisporre interfacce e procedure per un significativo ed alternativo intervento umano supplente sui sistemi ad alto rischio – che risponde proprio alla finalità di assicurare che la decisione ultima spetti ad una persona fisica che possa essere in ogni tempo chiamata a rispondere[25].

Il Regolamento predispone poi un regime sanzionatorio amministrativo (multe fino a 40 milioni di euro o percentuali del fatturato, sul modello GDPR) per le violazioni.

Sebbene il predetto provvedimento non introduca fattispecie penali, incide indirettamente sullo *ius puniendi*: in un senso, fissando standard tecnici e obblighi che potrebbero rilevare come regole cautelari nel giudizio sulla colpa penale (ad es. la mancata ottemperanza agli obblighi di trasparenza o *oversight* potrebbe costituire negligenza rilevante)[26]; nell'altro, obbligando gli Stati membri ad adeguare l'ordinamento interno, lasciando spazio a eventuali sanzioni penali nazionali in caso di usi illeciti dell'IA non coperti dal regime amministrativo[27].

DDL N. 1146/2024

In parallelo all'iniziativa europea, l'Italia ha varato la sua prima disciplina organica sull'IA. Il Disegno di Legge n. 1146/2024, approvato definitivamente il 17 settembre 2025, è una legge-quadro nazionale, *tabula rasa,* in materia di intelligenza artificiale[28].

Il *corpus* appena citato, è da integrarsi con l'AI Act, dovendosi porre sul binario di conformità del diritto UE, ma al contempo regolando aspetti specifici e nuovi da ritagliarsi sul contesto interno[29]. Il provvedimento delinea principi generali (trasparenza, proporzionalità, controllo umano, non

[24] *Ivi*
[25] O. Lombardi, "*Responsabilità penale dell'uomo per il danno cagionato attraverso condotte dolose e colpose nell'impiego dei sistemi di IA*", in Sistema Penale, 04/12/2024. <https://www.sistemapenale.it/pdf_contenuti/1736120025_lombardi-fasc-122024.pdf>
[26] *Ibidem*
[27] Si parla pertanto, di normativa *double hat*: in termini di efficacia diretta e di valore vincolante in termini normativi e valoriali per gli ordinamenti ricettivi.
[28] V. *Intelligenza Artificiale: approvata la legge delega italiana (ddl AI).* Op. cit.
[29] *Ibidem*

discriminazione, cybersicurezza ecc.[30]) e istituisce un'Autorità Nazionale per l'IA competente per vigilanza e attuazione (in coerenza con il modello del Regolamento UE).

Sul piano della rilevanza penale la legge introduce nuove fattispecie di reato e circostanze aggravanti per contrastare gli illeciti connessi all'uso di sistemi di IA[31].

In specifico:

1. Viene inserito nel Codice penale l'**art. 612-*quater* c.p**. – "*Illecita diffusione di contenuti generati o alterati con sistemi di IA*" che punisce con la reclusione da 1 a 5 anni chiunque, «*mediante l'impiego di sistemi di intelligenza artificiale*», falsifica o altera immagini, video o audio (c.d. *deepfake*) e li diffonda senza consenso cagionando un danno ingiusto alla persona rappresentata[32].

   Si tratta di una nuova figura criminosa volta a reprimere il fenomeno dei *deepfake* non consensuali, specie a contenuto sessuale, che la cronaca ha rivelato in preoccupante ascesa. Il reato è perseguibile a querela della vittima (salvo casi aggravati, ad es. se commesso contro persone incapaci o per finalità di intimidazione verso pubblici ufficiali). Questa disposizione colma un vuoto di tutela: finora, in assenza di norme *ad hoc*, casi di "porno-falsi" o video manipolati venivano infatti affrontati indirettamente e per il tramite di fattispecie spesso inadeguate, come la diffamazione o la violazione della *privacy*. La tipizzazione di 612-*quater* c.p. offre dunque uno strumento sanzionatorio mirato contro la diffusione di *fake* audio-video lesivi della reputazione, identità e dignità altrui[33].

2. Sono state introdotte **aggravanti specifiche** per reati già esistenti quando e se commessi mediante l'uso di sistemi di IA. Ad esempio, per la truffa (art. 640 c.p.); la frode informatica (art. 640-*ter* c.p.), i reati di riciclaggio e autoriciclaggio (artt. 648-*bis* e 648-*ter*.1 c.p.), ed altresì, taluni reati societari/finanziari (es. aggiotaggio informativo, art. 2637 c.c.) qualora risultino perpetrati e facilitati nella commissione da algoritmi tecnologici.

   L'uso dell'IA diviene in questi casi un elemento di speciale gravità: si presume che l'apporto artificiale ed artefatto dei più disparati ingegni possa aumentare la capacità offensiva del reato – ad esempio, rendendo più insidiosi gli schemi di frode online, potenziando campagne di *phishing* tramite *bot* intelligenti, o creando "*deepfake voice*" per ingannare le vittime (truffe del CEO impersonato, ecc.). Resta comunque ferma l'esigenza di bilanciare la repressione degli abusi tecnologici con la tutela di usi legittimi dell'IA – ad esempio distinguendosi nettamente i

[30] Legislatura 19ª - Disegno di legge n. 1146 | Senato della Repubblica < https://www.senato.it/show-doc?id=1418921&leg=19&tipodoc=DDLPRES&part=ddlpres_ddlpres1-articolato_articolato1 >
[31] V. *Intelligenza Artificiale: approvata la legge delega italiana (ddl AI).* Op. cit.
[32] Legislatura 19ª - Disegno di legge n. 1146 | Senato della Repubblica
[33] S. De Flammineis, op. cit.

deepfake satirici o artistici (esercizio di libertà di espressione) dalle falsificazioni ingannevoli e lesive[34].

3. Delega al Governo in materia penale: la legge 1146/2024 non esaurisce l'intervento punitivo con le norme sopra menzionate, ma affida al Governo una delega legislativa per integrare ulteriormente la disciplina penale dell'IA.

   In base all'art. 22, comma 5, della legge, dovranno essere emanati uno o più decreti legislativi volti a definire organicamente le fattispecie di illecito connesse a usi illeciti dell'IA[35]. I criteri direttivi della delega prevedono: (a) strumenti anche cautelari (in sede civile, amministrativa e penale) per inibire e rimuovere contenuti illeciti generati con IA, con sanzioni effettive e dissuasive; (b) l'introduzione di nuove fattispecie di reato, punite sia a titolo di dolo che di colpa, centrate sulla omessa adozione di misure di sicurezza adeguate nella produzione, immissione in commercio e uso professionale di sistemi di IA; (c) l'introduzione di ulteriori reati, puniti a titolo di dolo, a tutela di specifici beni giuridici che risultino esposti a rischi peculiari dall'uso di IA e non adeguatamente protetti dalle norme esistenti; (d) l'introduzione di una circostanza aggravante ad effetto speciale per i delitti dolosi (non puniti con ergastolo) in cui l'impiego dell'IA abbia inciso in maniera significativamente grave sull'offesa tipica, in particolare per reati contro la persona o contro lo Stato.

   Una delega, di ampissimo respiro, che mostra l'intento del legislatore di "anticipare" la tutela penale nel settore: non solo punire l'evento lesivo causato dall'IA, ma sanzionare anche (a monte) la mancata osservanza delle regole di diligenza e sicurezza nello sviluppo e utilizzo di tali tecnologie.

   Allo stesso tempo, il legislatore italiano dovrà coordinare queste innovazioni con il quadro UE: l'art. 22 richiama espressamente l'adeguamento al Regolamento europeo AI Act, evidenziando come la strategia nazionale debba restare complementare e coerente rispetto a quella comunitaria[36].

3. PROFILI DI IMPUTAZIONE PENALE: DOLO E COLPA, UTILIZZATORI VS. PRODUTTORI

Nel diritto penale vigente l'autore di reato può essere solo un soggetto umano, ed è la mancanza di libertà del volere che determina un *deficit* in senso di colpevolezza.

Gli eventi offensivi riconducibili alle innumerevoli forme di criminalità robotica, al personale senza anima, devono pertanto essere imputati – secondo i casi – all'utilizzatore (o operatore)

[34] Legislatura 19ª - Disegno di legge n. 1146 | Senato della Repubblica
[35] *Ivi*
[36] *Ibidem.* V. anche Intelligenza Artificiale: approvata la legge delega italiana (ddl AI). Op. cit.

dell'IA, ovvero al produttore (sviluppatore) della stessa, sempre che costoro abbiano agito con dolo o con colpa.

Occorre pertanto distinguere le possibili configurazioni di responsabilità penale legate ai c.d. 'attanti[37]' in base all'elemento soggettivo ed al ruolo del soggetto.

3.1 Responsabilità dell'utilizzatore e profili di responsabilità dolosa

L'utilizzatore di un sistema di IA (sia esso l'utente finale, il committente o il gestore operativo) risponde a titolo di dolo ogniqualvolta se ne serva come mezzo per commettere intenzionalmente un reato[38].

Potrebbe sembrare così risolta la questione del dolo, eppure, alcune questioni rimangono aperte e costituiscono oggetto di non meri esercizi intellettivi e retorici su cui parte della dottrina è ancora chiamata ad interrogarsi. Una fra queste: l'attribuzione al soggetto agente dei frutti della propria condotta, espletata con l'ausilio dei sistemi robotici, implica anche una piena ed approfondita comprensione dei sistemi di funzionamento (cosa che presupporrebbe una conoscenza tecnica-informatica superiore alla media), oppure è sufficiente essere consci di aver fatto uso dei sistemi di i.a[39]?

I principi tradizionali in materia di concorso di cause e di mezzi di esecuzione sono in grado di coprire tali situazioni: l'IA, anche se sofisticata, resta uno strumento nelle mani dell'uomo[40]: certamente non si ha a che fare con un 'Prometeo Scatenato' che sbarra al tentativo di accesso e di comprensione dall'esterno.

---

[37] Lett. *actant*, il prodotto frutto dell'ingegno di uno sperimentatore che si trova in situazioni subdole ed inaspettate, le quali lo definiscono attivamente. Così a p. 123, B. Latour, *Pandora's Hope,* Londra, 2000

[38] L'utilizzo massiccio di dispositivi intelligenti non finisce solo a supporto della vita e dei diritti dell'uomo, come *pars costruens*, vi è da considerare che la velocità evolutiva (in termini di precisione ed accuratezza di compiti ed attività svolte) trova corrispondente in una evoluzione di illeciti (sempre più precisi, difficili da decifrare) – *pars destruens*.
Non è un caso che i legislatori nazionali e sovranazionali debbano equipaggiarsi normativamente contro un nemico che a volte non è ben tratteggiato, e che comunque porta a conseguenze dannose per un bacino di utenti potenzialmente indefinito. Si v. ad es. la Convenzione del 24 dicembre2024, voluta dalle Nazioni Unite per la lotta al *cybercrime*.

[39] O. Lombardi, *Responsabilità Penale,* op. cit. v. p. 65 «*it has to be noted that understanding the workings of the AI system or its susceptibility to errors is necessary for the commission of an intentional crime in order to be able to draw conclusions about the causality from the perspective of the perpetrator and to hold the perpetrator responsible for intentional crime*»

[40] Non si tratta di una 'seconda natura', come molti hanno cercato di definirla. L'ingovernabilità che porterebbe a quello che Weber diceva essere 'disincantamento dal mondo' e che aprirebbe le porte ad una de-responsabilizzazione non è corretta, né può ascendere a giustifica per i prodotti umani nel reale. B. Fragrasso, *Intelligenza artificiale, op. cit.*, v. Pp. 70 e ss.

Esemplificativamente, chi impiega un algoritmo per perpetrare una frode informatica o per creare e diffondere consapevolmente notizie false lesive (c.d. *deepfake* volti a diffamare o a ricattare qualcuno) sarà responsabile del reato intenzionale corrispondente (truffa, diffamazione, estorsione, ecc.), con l'aggravante dell'uso di strumenti automatizzati quando prevista[41].

Il dolo copre in questo caso sia l'azione sull'IA (l'immissione di certi e specifici *input* o istruzioni con volontà criminosa) sia l'evento illecito prodotto grazie ad essa[42].

Un caso esemplare di dolo diretto è l'utilizzo di *deepfake* audio/video per commettere reati: ad esempio, generare falsi video pedopornografici integra il reato di produzione di materiale pedopornografico (come recentemente riconosciuto da un Tribunale spagnolo)[43]. Un esempio che mostra come l'intenzionalità nell'uso distorto dell'IA conduca ad una piena imputazione personale.

Un aspetto problematico concerne invece la ricostruzione e delimitazione del dolo eventuale in capo all'utilizzatore: situazioni in cui costui non miri direttamente all'evento lesivo, ma accetti consapevolmente un rischio elevato che l'IA possa provocarlo.

Si pensi all'operatore di un veicolo autonomo che, per battere un record di velocità, attivi una modalità non sicura sapendo che l'auto potrebbe causare incidente mortale: se tale evento si verificasse, potrebbe profilarsi a suo carico un dolo eventuale di omicidio, avendo egli accettato il rischio concreto di uccidere.

Tracciare il confine tra dolo eventuale e colpa cosciente nell'interazione uomo-macchina non è tuttavia semplice e richiederà l'apprezzamento fine e soppesato dei concetti di rischio conosciuto e voluto, rispetto al rischio sottovalutato[44].

In generale, l'utilizzatore risponde dolosamente di tutti gli esiti dannosi prevedibili e voluti (anche solo come possibilità accettata) derivanti dalla collaborazione uomo-IA[45].

---

[41] Legislatura 19ª - Disegno di legge n. 1146 | Senato della Repubblica

[42] Bisognerebbe, poi, interrogarsi sul significante di 'azione', e quindi se la definizione della prima sia quella del senso stretto e del comune ed ampio sapeere, considerati non potranno essere quei reati a forma vincolata, realizzati modalmente dal sistema robotico; al contrario, se concepire un'azione *lato sensu*, che comporterebbe sussumere al contegno umano tutte le attività tecnologiche, tanto da far rispondere l'utilizzatore anche per i reati a forma vincolata. B. Fragrasso, *Intelligenza artificiale*, op. cit., v. Pp. 554 e ss.

[43] *Spain: Court punishes schoolboys for spreading AI deepfakes of girls | Scottish Legal News* <https://www.scottishlegal.com/articles/spain-court-punishes-schoolboys-for-spreading-ai-deepfakes-of-girls>

[44] Schroeder attribuisce al rischio consentito una duplice valenza alternativa, quella cioè di rendere possibile l'esercizio di attività socialmente riconosciute, ed anche alleggerire l'attenzione, in situazioni imprevedibili e complesse, su pericoli «*speciali*». V. p. 251 di G. Forti, *Colpa ed evento nel diritto penale,* Milano, 1990

[45] O. Lombardi, "*Responsabilità penale…",* op. cit. v. Anche b. Fragrasso, *Intelligenza artificiale,* op. cit., pp. 520 e ss. Nel caso del macchinario che funge da *medium* fra l'uomo e l'evento finale e che non attenua né esclude il dolo se l'evento lesivo e il bene giuridico offeso rientravano nella rappresentazione e volontà dell'agente.

Diverso il caso in cui l'IA generi conseguenze ulteriori, non volute né previste dall'utente: in tali circostanze l'addebito a titolo di dolo è escluso per difetto dell'intenzione, potendosi configurare (al più) una responsabilità colposa (se l'utente avrebbe dovuto evitare l'evento con la dovuta diligenza) oppure nessuna responsabilità penale (se l'evento era imprevedibile e inevitabile, configurandosi, in tal caso, anche una depersonalizzazione del soggetto).

Se si optasse per sussumere l'attività artificiale a quella umana, bisognerebbe verificare in concreto che l'intera attività (incluso anche l'evento finale) posta in essere dalla tecnologia sia stata oggetto di rappresentazione e volizione da parte di chi agisce[46].

In dottrina si è posto in evidenza che, ai sensi dell'art. 43 c.p., il dolo richiede la rappresentazione e volontà in relazione ad uno *specifico fatto di reato* storicamente determinato[47]. Non è sufficiente un dolo "generico" o una volontà di massima di realizzare illeciti: se l'evento che occorre esula dalla sfera di ciò che l'agente si era rappresentato e voluto, viene meno l'elemento soggettivo doloso per quell'evento. Anche solo una deviazione dell'artificio strumentale che abbia innescato una causalità di eventi diversi rispetto a quelli intenzionalmente desiderati, comporterebbe uno scollamento fra il voluto e l'offesa che si è materializzata[48].

Esemplificando: un operatore lancia un algoritmo con l'intento di ottimizzare dei processi aziendali leciti, ma l'IA pone in essere attività illegali (fra molte, una violazione dei sistemi informatici di terzi o una discriminazione di candidati sul lavoro), l'intenzione criminosa dell'utente farebbe difetto; occorrerà valutare, caso per caso, se vi sia stata negligenza (il poter, al passato, accorgersi e impedire tali effetti) o se ci si trovi di fronte a un «r*esponsibility gap*»: un vuoto di responsabilità penale dovuto all'autonomia dell'IA[49].

---

[46]B. Fragrasso, *Intelligenza artificiale, op. cit.*, v. p. 547. Alcuni autori hanno provocatoriamente parlato di configurare l'IA come *"braccio"* o addirittura come coautore del reato, per poi imputare all'uomo un concorso ai sensi dell'art. 110 c.p. Come osservato, l'IA non è "umanizzabile" ai fini penalistici e, se le condotte verificatesi non erano prevedibili e volute dall'agente, non si può imputare a qualcuno un fatto che non ha voluto né previsto. In definitiva, la soluzione del concorso appare più metaforica che reale.

[47] *Ivi*, V. pp. 554 e ss. L'evento che integra il dolo sarà un fatto concreto, storicamente determinato e non un fatto generico analogo. Il dolo sarà escluso ai sensi dell'art. 47 c.p. solo quando vi sia un macroscopico scollamento fra il voluto e rappresentato dal lato umano, e quanto realizzato dalla macchina.

[48] *Ibidem.* V. Pp. 568 e ss. Si vedano i casi di *aberratio causae, ictus, delicti.*
Si dovrebbe inoltre aprire - considerato l'esito fuori scala, ed il rischio di un concreto vuoto di tutela - una digressione che prenderebbe forma da interrogativi come: '*e se si considerasse l'IA come esecutore materiale, o l'uomo come mandante?'.*
È di Consulich, l'idea di provare ad uscire dall'*impasse*, con l'*actio libera in causa*, che permetterebbe di muovere un rimprovero anticipato rispetto alla verificazione dell'offesa. Ed è lo stesso autore a tracciare i limiti alla sua stessa soluzione, essendo presenti nel Codice penale i limiti di applicazione normativa per questo istituto.

[49] S. De Flammineis, "*Fattispecie penali",* op. cit.

3.2 Responsabilità dell'utilizzatore per colpa

Nel mondo reale la gran parte delle interazioni quotidiane con sistemi di IA avviene senza finalità criminose. Fisiologicamente, si tende ad osservare un utilizzo da parte di persone o operatori che approcciano a strumenti intelligenti per scopi leciti (guidare un veicolo semi-autonomo, coadiuvare diagnosi mediche, gestire impianti industriali automatizzati, ecc.), strumentalmente per avere vita più facile.

In questi casi l'eventuale offesa a beni giuridici non deriva da un'intenzione criminale e patologica, ma da incidenti, malfunzionamenti o errori del sistema di IA, in concomitanza (spesso) con un uso non corretto o una vigilanza inadeguata da parte dell'uomo.

Qui il diritto penale è chiamato a misurarsi con la "collaborazione interattiva uomo-macchina": il primo, nello specifico, delega funzioni all'IA, ma rimane (o dovrebbe rimanere) in una posizione di controllo, di "contatto-asociale", rientrando al più nell'alveo della responsabilità colposa.

Si tratta di comprendere se i modelli attuali di attribuzione della responsabilità collimino con un fatto-reato scatenato da una *'opera autonoma'* di un algoritmo, solo assistito dalla presenza umana[50].

Detto in altri termini, al netto di un fatto storico già verificatosi con infinite e specifiche particolarità, non vi è chi non veda nella ricostruzione del nesso colpa-evento una difficoltà nel rintracciare i significanti giuridico-penali per l'imputazione colposa in capo all'agente[51].

L'interrogativo "chiave" è quali obblighi gravitino intorno all'utilizzatore al fine di evitare che l'IA cagioni danni e la risposta va cercata nei principi generali sulla colpa e in eventuali obblighi positivi specifici.

In linea generale, il fruitore di un dispositivo o programma avanzato è tenuto alla diligenza, prudenza e perizia esigibili in quello specifico contesto (art. 43 c.p.).

Se il sistema di IA è intrinsecamente pericoloso (ad es. un'auto a guida autonoma, un robot chirurgico, un qualsiasi macchinario industriale intelligente), l'utilizzatore assume il ruolo di garante della sicurezza: analogamente a quanto avviene per il conducente di un veicolo tradizionale o per il sorvegliante di una macchina, si avrà l'obbligo giuridico di impedire eventi offensivi derivanti dal funzionamento della cosa ai sensi dell'art. 40 cpv. c.p. (posizione di garanzia)[52].

---

[50] V. Manes, *L'oracolo algoritmo e la giustizia penale,* v. p. 2
[51] G. Forti, *Colpa ed evento nel diritto penale,* Milano, 1990
[52] O. Lombardi, "*Responsabilità penale…",* op. cit.

In altre parole, l'operatore non può affidarsi "ciecamente" all'IA: deve monitorarne l'operato e intervenire in caso di malfunzionamenti prevedibili.

Ad esempio, nel caso di veicoli a guida automatizzata di livello 3 (dove è previsto che il conducente sia pronto a riprendere il controllo nella immediatezza), se il "conducente-passaggero" omettesse di vigilare sulla strada e si verificasse un incidente mortale che si sarebbe potuto evitare agendo in tempo, si potrebbe contestare l'omicidio colposo per imprudenza/negligenza; il concetto usato, da un punto di vista normativo, è quello del c.d. *homo eiusdem*, il coscienzioso ed avveduto agente modello che nel circolo dei rapporti sociali avrebbe dovuto e potuto riconoscere il verificarsi di fatti dannosi.

Un caso paradigmatico dell'agente (modello) concreto appena descritto viene dagli USA: unico nel suo genere, l'incidente di Tempe (Arizona, 2018), è invero il primo caso di investimento mortale da parte di un'auto a guida autonoma.

La *safety driver* presente a bordo è stata ritenuta colpevole di condotta negligente, poiché distraendosi alla guida ha omesso di prevenire l'investimento della vittima[53]. In quell'episodio (*Uber self-driving car*), la conducente di *backup* si era messa a guardare un video sul cellulare omettendo di monitorare il veicolo; il giudice ha concluso che "*aveva un solo compito, tenere gli occhi sulla strada*", e la sua omissione è stata la *condicio sine qua non* del sinistro.

La conducente ha patteggiato una condanna per omicidio stradale colposo (*endangerment*) con tre anni di *probation*. Uber come azienda non è stata imputata in sede penale (le autorità hanno escluso profili di responsabilità penale *corporate*), ma l'indagine ha evidenziato che la società aveva disattivato il sistema automatico di frenata d'emergenza confidando esclusivamente sull'intervento umano[54].

Questo tragico esempio conferma che, finché le normative imporranno la presenza di un operatore umano (*human in the loop/on the loop*) per sistemi autonomi, la colpa ricadrà sul soggetto in caso di mancato impedimento di eventi prevedibili[55].

---

[53] Backup driver for self-driving Uber that killed Arizona pedestrian pleads guilty | Uber | The Guardian < https://www.theguardian.com/technology/2023/aug/01/uber-self-driving-arizona-deadly-crash>; < https://forbes.it/2020/09/17/uber-auto-a-guida-autonoma-uccide-pedone/ >

[54] *Ibidem.*

[55] B. Fragrasso, *Intelligenza artificiale, op. cit.*, v. Pp. 475 e ss.
Dovrebbe venire in aiuto, il concetto di *human oversight* (supervisore umano, *rectius* tecnico), un *wishful thinking* sorretto da una fragilità di impostazione: come si sorveglia ciò che è impossibile da controllare?
Lungi dall'impelagarsi in un mero artificio retorico, nel dire che l'utilizzatore sarebbe eletto a capro espiatorio di una catena umana lunga e ramificata, finendo per 'accollarsi' il peso di errori che potrebbero essere stati innescati da altri. Madeleine C. Elish ha provato a spiegare quanto appena detto, con un concetto figurato, quello della c.d. '*mis-attribution'* o *'crumple zone'*. Si fa riferimento (metaforicamente) al peso morale e giuridico che crollerebbe sul supervisore in ogni caso considerato. Il concetto-guida usato per profondere un insegnamento considera il mondo delle

Naturalmente, la colpa esige la violazione di una regola cautelare: occorre stabilire quali fossero in concreto gli obblighi dell'utente.

Ad esempio, l'AI Act – pur non rivolgendosi direttamente agli utilizzatori privati – richiede per certi sistemi un "controllo umano effettivo" (art. 14) e avverte del rischio di *overreliance*, eccessiva fiducia dell'utente nell'*output* della macchina[56].

Tali previsioni possono orientare l'interprete nazionale nel qualificare i doveri di diligenza: se una specifica applicazione (es. un *software* diagnostico medico) necessita per regolamento UE della supervisione umana, l'utente sanitario che la adoperi senza alcuna verifica critica potrebbe essere ritenuto colposamente responsabile di lesioni o (*in* extremis del) decesso del paziente, qualora il *software* commetta un errore grossolano e prevedibile (es. una diagnosi errata che il medico avrebbe dovuto correggere).

In generale, laddove esistono regole cautelari scritte[57], e queste restino violate, si cadrà nel caso di colpa specifica (art. 43 c.p.).

Un ambito dove quanto appena esplicitato si sta concretizzando è quella della sicurezza sul lavoro: l'uso di sistemi di IA in ambito industriale impone al datore di lavoro e ai preposti di aggiornare la valutazione dei rischi e di formare adeguatamente il personale; un infortunio cagionato da un robot privo di sistemi di arresto di emergenza potrebbe far emergere profili di colpa per violazione delle norme antinfortunistiche (D.Lgs. 81/2008).

Oltre alle regole scritte, vi è poi il campo delle regole cautelari non scritte, ovvero quello della comune prudenza esigibile.

Qui l'autonomia e opacità dell'IA complicano la possibilità dell'accertamento: dottrina e giurisprudenza si interrogano su quale grado di prevedibilità per un errore di IA sia ragionevole attendersi dall'uomo medio (o dall'utente medio esperto).

Se un sistema *machine learning* è per definizione "*unpredictable by design*" (imprevedibile nei singoli *output*, come nota la dottrina)[58], può l'operatore essere rimproverato per non aver prevenuto un esito lesivo raro e statisticamente imprevedibile?

La risposta tende a dipendere dall'analisi del rischio *ex ante*: se il tipo di evento era conosciuto o conoscibile (ad es. era già noto che l'auto poteva non riconoscere i pedoni in certe condizioni di luce), allora l'utente aveva un obbligo di cautela (non attivare l'auto in autonomia in quel

automobili, dotate di una zona di deformazione controllata, quella cioè che riunisce le componenti progettuali in grado di assorbire l'urto e salvare i soggetti presenti nell'abitacolo al momento dell'impatto.

[56] O. Lombardi, "*Responsabilità penale…",* op. cit.

[57] Fragrasso B., *intelligenza artificiale*, op. cit. pp. 490 e ss. v. Anche art. 13 e 14 AI Act.

[58] B. Fragasso, "*La responsabilità penale del produttore di sistemi di IA*", in Diritto Penale Contemporaneo, 13/06/2023 (scuola SSM Napoli) < https://www.sistemapenale.it/it/articolo/fragasso-la-responsabilita-penale-del-produttore-di-sistemi-di-intelligenza-artificiale >

frangente, o monitorare attentamente le condizioni di tempo e di luogo nel quale operava), e la sua omissione sarebbe colposa.

Se invece il fatto è originato da un comportamento anomalo, abnorme e totalmente inatteso dell'IA, collocato al di fuori di ogni ragionevole ed umana previsione, l'utente potrebbe andare esente da colpa (caso fortuito tecnologico).

Il diritto penale italiano già conosce situazioni analoghe: ad esempio, quella del conducente di veicolo non punibile se sia un guasto tecnico imprevedibile a provocare un incidente.

Su questo terreno si giocherà molto con la evoluzione giurisprudenziale: individuare il punto di equilibrio tra l'affidamento alla tecnologia (che di per sé non è vietato, specie se la tecnologia è avanzata e di supporto) e il dovere di controllo umano.

È prevedibile che in futuro standard tecnici ed eventualmente pronunce giurisprudenziali delineino il contenuto preciso della *diligentia diligentis* per l'utente di IA, magari mutuando criteri da settori affini (es. la regola dell'art. 590-*sexies* c.p. per la colpa medica, dove il medico che si attiene alle linee guida accreditate non è punibile per colpa lieve).

Non a caso, si discute se l'utilizzatore che segue fedelmente le istruzioni e i protocolli di sicurezza del sistema possa andare esente da responsabilità per gli esiti avversi "residuali": è il concetto della "*residual risk*" accettabile, previsto anche dall'AI Act (art. 9 par.4) in ambito di conformità tecnica[59].

In altre parole, se la tecnologia complessivamente riduce il rischio rispetto all'azione umana media, gli incidenti rari ed inevitabili non dovrebbero dar luogo a sanzione penale: tema su cui si tornerà parlando delle proposte *de iure condendo*[60].

3.3 Responsabilità del produttore/sviluppatore

L'altro grande attore in scena, nel teatro della intelligenza artificiale, è il produttore o progettista degli stessi sistemi oggetto di trattazione.

Il primo opera "a monte", nella fase di creazione e immissione sul mercato del sistema, e potrebbe, in virtù della posizione lavorativa che ricopre, sembrare distante dall'eventuale illecito commesso "a valle" quando il sistema è già nelle mani di qualcun altro e le interazioni produttore-macchina non sono più una opzione considerabile.

[59] O. Lombardi, "*Responsabilità penale…",* op. cit.
[60] *Guida autonoma e responsabilità penale: serve rischiare*. - Lexit < https://www.lexit.it/guida-autonoma-e-responsabilita-penale-serve-rischiare/ >

Ciò non toglie, che è proprio l'apporto creativo iniziale a permettere la configurazione delle specifiche del sistema di IA considerato (i limiti di sicurezza, i meccanismi di controllo, ecc.), ed è per questo motivo che la riflessione penalistica ha indagato su eventi lesivi innescati da malfunzionamento da algoritmo; specificatamente sui profili della colpa, tipicamente per negligenza, imprudenza o imperizia nella progettazione o per la messa in commercio.

Nel contesto IA, questo schema viene sollecitato da nuovi problemi: il produttore crea un sistema dotato di una autonomia decisionale che può provocare esiti non anticipati o preventivati neppure dal programmatore stesso. Quanto rende complessa sia la prova del nesso causale tra una scelta progettuale e l'evento lesivo, sia la valutazione della colpa del produttore, tanto più allorquando l'evento era "concretamente imprevedibile" *ex ante*[61].

La dottrina si esprime in termini di crisi del modello nomologico-deduttivo del nesso causale e della colpa: in assenza di leggi scientifiche affidabili che spieghino il comportamento dell'IA (il cui apprendimento è spesso sub-ottimale e opaco[62]), il giudizio controfattuale causale si fa ipotetico, e imputare un errore "misterioso" al produttore rischia di tradursi in un'aleatoria responsabilità oggettiva postuma[63]. Il che non significa, tuttavia, escludere sempre la responsabilità del produttore. Al contrario, se il danno derivasse da carenze progettuali o produttive note o evitabili secondo lo stato dell'arte, il produttore dovrebbe rispondere.

In dottrina si sottolinea che *gli standard tecnici fungono da doppio argine*: da un lato, integrarli come regole cautelari scritte e dettagliate, consente di evitare la "deriva oggettivistica" (punire il produttore anche se ha seguito lo stato dell'arte: il caso in cui cioè ha fatto tutto il possibile per evitare danni o pericoli); dall'altro, essi forniscono al giudice parametri per valutare concretamente la prevedibilità e prevenibilità dell'evento[64].

---

[61] B. Fragasso, "L*a responsabilità penale",* op. cit. v. P. 56. Esemplificativamente, un'azienda sviluppa un algoritmo di navigazione per auto autonome; dopo migliaia di ore di funzionamento corretto, in una situazione peculiare, il *software* prende una decisione fatale (sterza improvvisamente causando un impatto). Se tale scenario non era mai stato rilevato nei test e non esisteva conoscenza scientifica del *bug*, attribuire colpa al team di sviluppo risulta arduo: manca la violazione cosciente di una regola cautelare, versandosi piuttosto nell'ambito del rischio tecnologico intrinseco (*danger inherent out not manageable*).

[62] *Ibidem v. Pp. 380 e ss.* C.d. *black box causality*, è chiaro che la causa dell'evento sia da correlarsi ad un apporto algoritimico, quello che rimane opaco è il perché del danno. Un esempio è atto a spiegare meglio quello che sopra è esplicitato solo a livello teorico. Si immagini la morte di un uomo causata da un veicolo a guida autonoma; le leggi scientifiche non possono spiegare il motivo di un evento acclarato: morte dell'uomo. L'evidenza fattuale ci consentirebbe solo di dire che causalmente, soltanto la non immissione sul mercato del veicolo avrebbe impedito la l'evento-morte.

[63] B. Fragasso, "L*a responsabilità penale",* op. cit.

[64] Il rischio di studiare, ma non comprendere sempre appieno la prevedibilità dei pericoli correlati alle interazioni dei sistemi di i.a., lascerebbe spazio di manovra al giudice che si renderebbe produttore di una regola cucita posteriormente sul modello di agente, *ad hoc,* che posteriormente avrebbe dovuto comportarsi in maniera diversa (la differenza dibattuta in dottrina fra agente modello ed agente in concreto).

Esemplificativamente, se un costruttore di robot chirurgici omettesse volutamente i dovuti collaudi o ignorasse problemi noti di sicurezza, al solo fine di accelerare l'immissione sul mercato, ed un paziente morisse per un malfunzionamento già segnalato, tale costruttore potrebbe essere incriminato per omicidio colposo (avendo violato le regole cautelari di settore).

In Italia si sono già avuti procedimenti penali per produttori di beni difettosi: si pensi ai processi per veicoli con difetti di fabbricazione letali, o per dispositivi medici non conformi.

Nel caso IA a fare la differenza sarà la normazione tecnica: la presenza di standard ISO o normative UE che prescrivono requisiti di sicurezza per algoritmi (es. obbligo di *risk assessment,* di *dataset* non discriminatori, di *fail-safe* in caso di errore) potranno fondare la colpa del produttore che non li rispetti.

Da siffatta prospettiva, le previsioni dell'AI Act sui fornitori (ed i relativi obblighi di assicurare qualità dei dati, documentazione tecnica, registri di eventi, ecc.) potrebbero diventare parametri di diligenza professionale: la violazione grave di quelle prescrizioni– se sfociasse in un disastro – costituirebbe un forte elemento accusatorio in sede penale[65].

Per il tramite della delega contenuta nel DDL 1146/2024, l'ordinamento italiano potrebbe inoltre introdurre un'apposita fattispecie di omessa adozione di misure di sicurezza nella produzione e messa in commercio di IA punita anche a titolo di colpa[66].

Una siffatta norma configurerebbe un reato di pericolo, non si può pertanto non trattare di alcuni nodi critici rilevati. Primo fra questi è una sorta di sovrapposizione fra condotta omissiva e commissiva; l'ideatore risponderà non per aver commesso il fatto lesivo direttamente (*agere*), ma per aver omesso di prevenirlo. In altri termini, per non aver implementato cautele adeguate, prescindendosi dunque dal fatto che l'evento si verifichi o meno (similmente a quanto già avviene per le contravvenzioni in materia di sicurezza del lavoro).

Ciò consentirebbe di anticipare la tutela penale, sanzionando il produttore negligente già prima che un incidente accada e non perché formalmente egli abbia un obbligo cautelare scritto; il rimprovero è quello di una colpa omissiva, nel libero svolgimento di un'attività lecita pericolosa senza le dovute cautele. In tal senso, il confine tra azione e omissione sfuma: l'azione colposa è immettere sul mercato un prodotto difettoso/insicuro, ciò implica al contempo l'omissione di misure dovute.

[65] O. Lombardi, "*Responsabilità penale…",* op. cit.
[66] Legislatura 19ª - Disegno di legge n. 1146 | Senato della Repubblica

Spetterà poi al giudice non confondere i piani, non richiedere al produttore l'impossibile, limitandosi al governabile e quindi a ciò che rientra nell'ordinaria sfera di controllo di chi progetta[67].

Un cenno va ancora fatto anche alla responsabilità penale delle persone giuridiche (D.Lgs. 231/2001). Se un reato (es. omicidio colposo) commesso nell'utilizzo di un'IA risultasse imputabile a un dirigente o dipendente di un'azienda nell'interesse o a vantaggio di quest'ultima, la società potrebbe essere chiamata a rispondere in sede amministrativa ex 231.

Ad esempio, una casa automobilistica che, per risparmiare, omettesse intenzionalmente test di sicurezza sull'AI di guida potrebbe, in caso di incidente mortale, essere passibile di sanzione *ex* D.lgs 231/2001 qualora il fatto risultasse riconducibile ad una fattispecie presupposto (omicidio colposo per violazione di norme antinfortunistiche rientra nel catalogo).

Va ricordato che il modello 231 richiede comunque un reato commesso da persona fisica colpevole: non risolvendosi il caso in cui nessun individuo sia penalmente colpevole (il *responsibility gap*). In altre parole, se un evento dannoso da IA non sia imputabile per dolo o colpa a nessun soggetto, nemmeno l'ente collettivo potrebbe essere sanzionato, difettando il reato presupposto. Conclusione che rafforza l'idea che per le IA più "autonome" occorra costruire un sistema di regole e responsabilità ad hoc, per evitarsi zone franche[68].

3.4 Cenni comparativi e problematiche generali

In ordinamenti stranieri si sono esplorati approcci differenti.

Nel *common law*, ad esempio, si discute se attribuire la responsabilità del danno da IA al '*culpable programmer*' (programmatore colpevole) o al c.d. '*end-user*' a seconda dei casi, oppure se

[67] B. Fragrasso, *Intelligenza artificiale,* op, cit., v. Pp. 460 e ss.
Il produttore non può rispondere per non essere stato onnisciente, né si può ammettere una colpa che sussuma a sé una classe di eventi iper-generici, così da avere una vera 'taglia' sempre pendente da espiare seguendo un unico comandamento: '*in dubio pro omittere*'.
L'obiettivo del legislatore non è quello di evitare il verificarsi di ogni evento lesivo, si tengano sempre a mente le solide certezze del mondo robotico: l'opacità delle macchine e la efficacia preventiva delle regole cautelari - in relazione a classi di sotto-eventi pericolosi e più o meno dettagliati (non vi può essere certezza dogmatica fra *input* X e *output* Y) - apprezzabili solo in ottica complessiva.

[68] V. Mongillo, *Responsabilità da reato degli enti e crimini connessi all'intelligenza artificiale: tecniche giuridiche di intervento e principali ostacoli.* L'autore evidenzia come pene rivolte direttamente ai sistemi di IA, e le sanzioni rivolte agli enti, non trovino adeguati basi considerato lo stadio attuale delle conoscenze scientifiche. Attualmente per quanto concerne la *corporate governance* le alternative più accreditate sono due: i) colpire l'organizzazione che non abbia adottato misure di controllo idonee al contenimento del rischio per i risultati avversi da IA; ii) colpire la *societas* che commercia e produce prodotti digitali, omettendo di intraprendere misure di *compliance* legale e *standard* prestabiliti *ex lege*. v. P. 25

utilizzare dottrine come la *vicarious liability* (responsabilità indiretta di un superiore per l'atto altrui) in analogia con la responsabilità del padrone per l'animale.

Nessuna giurisdizione democraticamente evoluta ha però sinora riconosciuto personalità giuridica penale a un sistema di IA: un tentativo concettuale in tal senso (l'idea di creare una sorta di '*electronic person'* per i *robot* avanzati) fu avanzato peraltro in sede UE per la responsabilità civile, ma poi fortemente respinto per quella penale[69].

Lo scenario attuale vede dunque confermato il principio: "l'uomo al centro". L'IA è fonte di nuove insidie, ma il diritto penale vi fa fronte con gli strumenti classici dell'imputazione umana: dolo per gli usi criminosi consapevoli, colpa (anche specifica) per le omissioni o negligenze sia dell'utilizzatore che del produttore.

Resta il margine di casi in cui nessuno appare colpevole (guasti imprevedibili, errori spontanei del sistema non evitabili): in tali evenienze potrà parlarsi di evento fortuito non punibile – per quanto insoddisfacente possa apparire dal punto di vista emotivo - resta invero preferibile ad artificiose estensioni della colpevolezza in assenza di suoi presupposti. Sarà il legislatore a dover (semmai) valutare i correttivi futuri, di cui appresso.

4. L'INTELLIGENZA ARTIFICIALE ED I PALAZZI DI GIUSTIZIA

Anche gli attori nelle aule di giustizia devono prendere atto che l'uso della intelligenza artificiale li riguarda e li tange da vicino, non solo come regolatori degli aspetti che hanno ad oggetto la vita dei terzi, ma per regolarizzare, acquisire e riordinare informazioni di tipo giudiziario e giurisprudenziale al fine di produrre *outputs* risolutivi di singole questioni o di intere decisioni[70].

Leibniz non sarebbe rimasto sorpreso di come si tenda oggi ad appiattire le discussioni dei primi a *calamacos et abacos* e a far convertire oratori (giudici, avvocati) in meri *calculatores*.

In questi termini l'utilizzo di c.d. *decision making* che soppiantano gli autori ed attori umani della giurisdizione, portano a dover sviscerare il tema dei modelli di AI «*forti».*

Parrebbe però che soltanto l'uso della IA nelle Corti possa far rivivere il giudice per come lo intendeva Cesare Beccaria[71], il persecutore della messa al bando dello spirito della legge per far trionfare il sillogismo perfetto, mero prodotto di calcoli matematici.

---

[69] S. De Flammineis, "*Fattispecie penali",* op. cit.

[70] G. Canzio & L. Lupària Donati (a cura di), *Prova scientifica e processo penale,* 2025. V. pp. 999 e ss.

[71] C. BECCARIA, Dei delitti e delle pene (1764), Milano, 1981, § IV, v. p. 16. «*In ogni delitto si deve fare dal giudice un sillogismo perfetto: la maggiore dev'essere la legge generale, la minore l'azione conforme o no alla legge, la conseguenza la libertà o la pena. Quando il giudice sia costretto, o voglia fare anche soli due sillogismi, si apre la porta all'incertezza*».

Gli algoritmi tecnologici che impattano sul mondo Giustizia sono (ai sensi dell'art. 6 par. 2 del Reg. UE) sistemi ad alto rischio e per tale motivo si è optato per una non limitazione della continua ed alimentata interfaccia uomo-macchina: un contatto sociale e necessario (insomma) che consenta al *deployer* di interpretare e valutare l'esternalizzazione delle macchine in termini di impatto e messa in pericolo dei diritti fondamentali[72].

Per la prima volta anche internamente, il legislatore deve prepararsi ad affrontare una sfida inedita, quella fra giurisdizione e sistemi artificiali in un mondo (quello del diritto penale) che è di fatto e necessariamente plasmato da e per l'uomo, fallibile anche nelle decisioni non esatte, ma più giuste.

Lo scenario polarizzato che sembrerebbe aprirsi al tecnico del diritto è quello di optare fra tecnologia e tecnocrazia e quindi la scelta fra il mondo del sapere e delle operazioni affidate al giudice ed alle parti e le evidenze probatorie frutto di un calcolo algoritmico, probabilistico e disumanizzato.

Gli scenari di una giustizia erronea, arbitraria, che dipende dal "se il giudice abbia fatto una buona colazione" (il c.d. *breakfast sentencing*), i mille dubbi che celano trappole cognitive e scotti da pagare per i terzi, il raziocinio limitato di un essere pensante, l'incidenza dell'umore e dei "rumori quotidiani" su ogni valutazione e decisione che quotidianamente compiamo, ci potrebbero tentare nella sollecitazione di una via più rapida: quella dei correttivi artificiali che consentano di migliorare in termini di affidabilità e "calcolabilità" la decisione giudiziale[73].

Attualmente il quesito non è sull'*an* dell'utilizzo potenziale della tecnologia nelle fasi procedimentali, ma è sul come e cosa utilizzare dei sistemi tecnologici conosciuti e governabili, rischiandosi di mettere a repentaglio la liturgia della legge e della parola. Nella terra del dover stare al passo con i tempi si è approdati a due opzioni:

- gli algoritmi come strumento di supporto nella ricerca della prova[74]

---

[72] G. Canzio & L. Lupària Donati (a cura di), *Prova scientifica e processo penale,* op. cit.*,* v. p. 1008.Cuscinetto protettivo ai sensi degli artt. 85 ed 86 (AI ACT) è la previsione di presentare reclamo qualora ci si accorga della violazione del suddetto regolamento e dall'altro la possibilità di chi subisce una decisione adottata dall'utilizzatore dei sistemi algoritmici, capaci di ledere ed incidere sulla salute, sicurezza umana e diritti fondamentali, di chiedere ed ottenere dal primo, spiegazioni e giustificazioni sulla procedura e sui principali elementi che sorreggono la decisione adottata.

[73] R. Perona, *ChatGPT e decisione giudiziale: per un primo commento alla recente sentenza del Juzgado Primero Laboral di Cartagena de Indias (Colombia),* < *ChatGPT e decisione giudiziale: per un primo commento alla recente sentenza del Juzgado Primero Laboral di Cartagena de Indias (Colombia) - Diritti Comparati* >. Nel febbraio 2023, il giudice Padilla di Cartagena, in Colombia, ha emesso una decisione basata su una risposta generata da un programma di intelligenza artificiale, il chatbot GPT. Il giudice ha posto una domanda al sistema riguardante una questione fiscale, specificamente sulle detrazioni fiscali per le cure mediche di un portatore di handicap. La risposta fornita dall'IA è stata considerata convincente e, successivamente, è stata utilizzata come base per redigere la sentenza.

[74] G. Canzio & L. Lupària Donati (a cura di), *Prova scientifica, op. cit.*, v. p. 1029

- l'utilizzo dei c.d. *risk assessment tools:* i mezzi di prova che attengono alla giustizia predittiva e al supporto al giudice nella valutazione del rischio[75].

Sembrerebbe quindi *bypassato* il sentimento di diffidenza che circondava i rischi e le opportunità dell'uso e l'impiego della c.d. *hybrid IA*, avendosi ben saldo che la macchina non deve avere una efficienza euristica e che i rischi di un cieco affidamento a decisioni ottenute con calcoli probabilistici celerebbero vizi occulti in grado di incidere sugli individui esponendoli a pericoli che tangerebbero i loro diritti, le loro garanzie e la libertà personale. Un dominio assoluto ed *absoluto* della macchina sull'uomo.

Almeno nelle aule dei tribunali e negli studi legali la considerazione che si ha della tecnologia è pertanto quella di un impiego, potenzialmente fruttuoso, ma pur sempre strumentale, informativo ed a supporto dell'unica decisione ammissibile: quella (seppur a volte fallibile), ma umana e ragionata[76].

5. GIURISPRUDENZA RECENTE, I CASI PRATICI DELLA IA E LA RILEVANZA PENALE

Negli ultimi anni stanno emergendo i primi casi concreti che, impattando sul reale, mettono alla prova i principi dello *ius puniendi* nell'era dell'IA, sia in Italia che all'estero in ambito comparato europeo.

Di seguito alcuni esempi di vicende giudiziarie significative (2022-2025):

- veicoli a guida autonoma e incidenti stradali: l'Italia si prepara a testare su larga scala veicoli con pilota automatico, ma non ha ancora registrato sentenze definitive su incidenti causati da essi.

  Il precedente più vicino proviene dagli Stati Uniti ed è quello citato di Uber (Tempe, 2018). In Europa, l'attenzione è alta: la Germania ha introdotto norme sulla circolazione di veicoli autonomi già nel 2017, imponendo per legge che il sistema rimanga sotto supervisione umana e che in caso di malfunzionamento il conducente riprenda immediatamente il controllo.

  Nel caso finora ipotetico di un sinistro mortale in modalità autonoma, la dottrina tedesca prospetta (anch'essa) l'incriminazione del conducente (se ha tardato a intervenire), oppure del produttore (se l'incidente sia dovuto a un difetto tecnico prevedibile), escludendo la punibilità dell'IA in sé.

[75] *Ibidem*, v. p. 1037
[76] G. Canzio & L. Lupària Donati (a cura di), *Prova scientifica e processo penale,* op. cit. v. pp. 10433 e ss.

In Italia, possiamo citare una pronuncia del GIP di Milano (caso inedito, 2023[77]) che, con riferimento ad un incidente con auto dotata di sistemi ADAS avanzati, ha archiviato il procedimento nei confronti del conducente ritenendo che questi avesse adottato tutte le cautele e che l'errore del *software* fosse totalmente imprevedibile: un segnale di come i giudici potrebbero riconoscere il caso fortuito tecnologico.

- *deepfake* e reati contro la persona: il fenomeno dei *deepfake* (video/foto/audio falsificati tramite IA) ha già prodotto interventi normativi – come detto, l'Italia, con l'art. 612-*quater* c.p. di recente introduzione – ma anche alcuni casi giudiziari rilevanti.

  Oltre al caso spagnolo di Almendralejo, merita ricordarsi che i casi nostrani più in voga sono di indagini di *deepfake* a scopo di *revenge porn* e diffusione non consensuale di immagini intime. Ad esempio, nel 2023 la Procura di Bologna ha indagato un giovane accusato di aver utilizzato un'app di *deepfake porn* per creare falsi nudi di ex compagne di scuola e averli diffusi su *Telegram*[78]: in assenza (a quell'epoca) di una specifica fattispecie si è fatto ricorso alle disposizioni che regolavano la diffamazione aggravata e la produzione di materiale osceno, con il sequestro preventivo degli *account*. Questo caso ha fatto da "apripista" per la sensibilizzazione sul tema, influenzando la scelta del legislatore di intervenire col nuovo 612-*quater* c.p.

  A livello europeo, oltre alla Spagna, segnaliamo nel Regno Unito un provvedimento peculiare: nel 2023 la High Court di Londra ha emesso un'ingiunzione ("*deepfake pornography injunction*") per proibire la diffusione di video falsi pornografici raffiguranti una nota influencer, riconoscendo così giurisdizionalmente la lesività dei deepfake anche in sede civile (tutela della privacy e *dall'harassment*); sul piano penale, il Regno Unito sta introducendo nel Online Safety Bill un reato specifico analogo al 612-*quater* c.p. italiano, a riprova di una convergenza internazionale nel reprimere penalmente tali condotte lesive della dignità personale.

- IA e crimini informatici/predittivi: già oggi la giurisprudenza affronta casi di *cybercrime* potenziati dall'IA. La Polizia Postale italiana, nelle sue relazioni annuali, ha documentato nuovi schemi di *phishing* e di *social engineering* in cui *bot* basati su IA generativa (*chatbot* avanzati) interagiscono con le vittime simulando conversazioni realistiche per carpire dati o denaro.

  Quando tali truffe vengono scoperte, i responsabili vengono perseguiti per i reati tradizionali (truffa, accesso abusivo a sistemi, frode informatica); l'aspetto "IA" emerge come elemento di fatto che può incidere sulla gravità (ad es. la condotta è più subdola e

[77] https://www.ilsole24ore.com/art/se-veicolo-ha-adas-costruttore-responsabile-salvo-colpe-guidatore-AERoE9qD?refresh_ce=1

[78] Pp. 588 e ss. V. il caso del CEO di Telegram, Pavel Durov, che fu arrestato nel 2024 per la commissione di una serie di reati verificatisi sulla piattaforma ad opera di utenti che scambiavano e riproducevano materiale pedopornografico.

seriale se automatizzata) e dunque sul trattamento sanzionatorio, ma senza aver (finora) prodotto pronunce di principio innovative.

Sul fronte delle IA predittive applicate al contrasto del crimine, va citata la recente giurisprudenza amministrativa e costituzionale: ad esempio, il Consiglio di Stato italiano, con sentenza n. 881/2020, ha annullato l'uso sperimentale del software SARI (riconoscimento facciale in tempo reale) da parte del Ministero dell'Interno, per violazione della normativa privacy e dei principi di proporzionalità, ravvisando il rischio di ingiuste segnalazioni e profilazioni di massa. Sebbene ciò esuli dal processo penale in senso stretto, incide sulla fase investigativa penale: l'utilizzo di algoritmi per individuare sospetti è guardato con estrema cautela, in quanto potenzialmente lesivo di diritti fondamentali e fonte di errori giudiziari (falsi positivi). A livello europeo, un caso emblematico è stato quello del sistema SyRI olandese: pur trattandosi di prevenzione delle frodi in ambito sociale, la sua "bocciatura" in tribunale ha creato un precedente che scoraggia applicazioni di data *mining* indiscriminato per fini di *law enforcement*.

Nel Regno Unito, nel caso Bridges v. South Wales Police (2020), la Court of Appeal ha dichiarato illegittimo l'uso di telecamere con riconoscimento facciale ai fini di sicurezza pubblica citando la mancanza di garanzie sufficienti contro arbitri e *bias* razziali: un chiaro monito in ordine ai limiti dell'uso di IA in ambito preventivo[79].

In sintesi, le sentenze recenti in Europa trasmettono un messaggio: sì all'uso di IA in ambito penale come strumento, ma a condizione che vi sia rispetto rigoroso dei diritti e che la responsabilità ultima resti in capo a operatori umani identificabili. Il sistema penale deve dunque muoversi tra due esigenze: cogliere i benefici dell'IA (ad es. strumenti investigativi più efficaci, analisi predittiva per allocare risorse di polizia) e scongiurare che l'IA mini le basi del giusto processo o attribuisca colpe senza volto.

- IA e processo penale: a margine, va ricordato che la giustizia penale stessa inizia a confrontarsi con l'IA come supporto decisionale. In alcune giurisdizioni (USA) si sono impiegati algoritmi di *sentencing* o *risk assessment* (es. algoritmo COMPAS per valutare rischio recidiva); ciò che ha suscitato critiche e ricorsi per la scarsa trasparenza e possibili discriminazioni[80].

[79] Nel caso Bridges, il signor Edward Bridges aveva contestato l'utilizzo da parte della South Wales Police del sistema di riconoscimento facciale in tempo reale ("*live biometric facial recognition* – LBFR") negli spazi pubblici. https://www.biometricupdate.com/202008/live-biometric-facial-recognition-use-by-south-wales-police-ruled-unlawful-by-appeals-court?utm_source=chatgpt.com

[80] L. Macrì, "*I primi passi dell'Italia verso l'impiego dell'IA nel processo penale*", in Giurisprudenza Penale Web, 2/2025 (sulle implicazioni di IA in ambito processuale) < https://www.giurisprudenzapenale.com/2025/02/06/i-primi-passi-dellitalia-verso-limpiego-dellia-nel-processo-penale-e-il-calcolo-del-rischio-di-recidiva/>

In Italia l'IA nel processo penale è, per ora, limitata a progetti pilota (es. programmi ministeriali per smistare fascicoli o cercare giurisprudenza). La dottrina e la magistratura di vertice (cfr. relazione Primo Presidente Cassazione 2020) hanno, tuttavia, già posto in guardia: qualsiasi uso di IA nel giudicare deve rispettare il "significativo controllo umano" e non violare l'obbligo costituzionale di motivazione del giudice[81].

5. PROPOSTE DE IURE CONDENDO

Di fronte ai vuoti normativi e ai nuovi dilemmi posti dall'IA, la dottrina penalistica italiana ha elaborato diverse proposte *de iure condendo* (riforme auspicabili) per adeguare il sistema penale ai cambiamenti estemporanei, ma senza tradirne i principi.

Alcune di queste idee hanno già influenzato il legislatore (si veda la Legge delega 1146/2024), altre restano spunti futuribili. Di seguito, le principali.

Esplicita esclusione della responsabilità penale "diretta" delle IA: più che una proposta, è una presa di posizione condivisa dalla quasi totalità degli studiosi: il divieto di attribuire soggettività penale agli agenti artificiali va mantenuto fermo, perché radicato nell'art. 27 Cost[82].

Si è osservato che una macchina non può essere imputata poiché priva di libero arbitrio e incapace di dolo o colpa in senso giuridico[83]. La pena verrebbe inoltre svuotata del senso rieducativo (impossibile "rieducare" un *software*)[84].

Questa apparente ovvietà è stata ribadita anche per contrastare le suggestioni provenienti dall'estero, come la teoria dell'autore israeliano Gabriel Hallevy[85] che ipotizzava tre modelli per imputare crimini alle IA (per analogia con la responsabilità del datore per il dipendente, o del proprietario per l'animale)[86].

[81] *Ibidem*
[82] S. De Flammineis, "*Fattispecie penali",* op. cit
[83] B. Fragrasso , *Intelligenza artificiale*…, op. cit. p. 47 e ss.
[84] S. De Flammineis, "*Fattispecie penali",* op. cit
[85] G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systams,* Springer, 2015. È uno studioso isreliano, principale teorizzatore della dovuta responsabilità diretta in capo ai sistemi di i.a.
[86] Il timore dell'*horror vacui* nei termini di tutela del penale ha portato alla elaborazione di "e-person" (persona elettronica); si legge, al punto 59 lettera f) della **risoluzione Resolution P8_TA(2017)0051 – Civil Law Rules on Robotics del European Parliament** (16 feb 2017): «*creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently*».

La dottrina italiana definirebbe quel dibattito '*much ado about nothing*[87]', riaffermando che '*machina delinquere non potest'* non è solo massima storica, ma vincolo costituzionale. In ambito morale e filosofico si suole dire che responsabilità implica controllo. In ambito giuridico-penale ciò si traduce nello scrutinare la posizione di garanzia e i doveri dell'uomo rispetto alla macchina[88].

Si è proposto di chiarire (semmai) legislativamente che l'IA vada trattata alla stregua di mera cosa ai fini dell'art. 40 cpv. c.p., così da evitarsi in radice fantasiose teorie sull'IA come soggetto attivo di reato.

Introduzione di reati omissivi specifici e "posizioni di garanzia" ex lege: come visto, la legge delega già spinge in questa direzione[89].

Diversi autori[90] suggeriscono di normare espressamente gli obblighi di chi utilizza o immette in commercio IA. Ciò significherebbe introdurre in legge o in regolamento (ad esempio) l'obbligo per il *deployer* di un'IA ad alto rischio di predisporre misure di sicurezza e monitoraggio continuo, qualificando tale soggetto come garante per la prevenzione di certi eventi (analogamente a come il datore di lavoro è già garante della sicurezza dei lavoratori)[91].

L'inadempimento di tali obblighi potrebbe essere sanzionato con una contravvenzione *ad hoc* (punita anche in assenza di evento) e, in caso di verificazione di un disastro, servirebbe a fondare la colpa specifica nell'omicidio o lesioni colpose.

In questo senso, dottrina e prassi applicativa, potrebbero mutuare schemi dal diritto penale della sicurezza sul lavoro o dell'ambiente, settori nei quali si è passati da un approccio reattivo (punire solo il fatto lesivo) ad uno preventivo (punire anche l'omessa cautela che crea situazioni pericolose).

---

[87] M. Cosulich, "*Much Ado About Nothing", ovvero dell'inutile compressione dell'eguaglianza del voto nella vigente legislazione elettorale parlamentare, <Cosulich.pdf >*

[88] O. Lombardi, "*Responsabilità penale…",* op. cit.
È oramai notizia di qualche tempo fa, la creazione negli Usa di un software (Do *not pay*), pronto e predisposto a fare le veci di avvocati. La startup fondata nel 2015 da Joshua Browder si poneva l'ambizioso obiettivo di automatizzare l'assistenza legale (contestare multe per parcheggio, offrire strumenti per risolvere dispute di modesta entità, reclami per voli in ritardo e cancellazioni di abbonamenti). Ad oggi però l'applicazione rimane nell'etere e la capacità di sostituire gli avvocati rimane una possibilità (e per le più varie e disparate ragioni) molto remota.

[89] Legislatura 19ª - Disegno di legge n. 1146 | Senato della Repubblica

[90] O. Lombardi, "*Responsabilità penale…'',* op. cit. e P. Troncone, *Il sistema dell'intelligenza artificiale nella trama grammaticale del diritto penale. Dalla responsabilità umana alla responsabilità delle macchine pensanti: un inatteso return trip effect, Cass. Pen. 09/22 < Troncone Intelligenza artificiale Cass. pen. 2022.pdf >*

[91] O. Lombardi, "*Responsabilità penale…",* op. cit.

"*Safe harbor*" per l'uso di IA sicure – la provocatoria proposta Lamesso.

Artefice di una delle idee più innovative, quella avanzata dall'avv. Federico Lamesso [92], capovolgerebbe le prospettive: prevedere una sorta di esenzione da responsabilità penale nei (pochi) casi in cui la tecnologia IA abbia dimostrato un'affidabilità statistica superiore a quella umana.

Si tratterebbe, in sostanza, di riconoscere che, se un sistema autonomo commette meno errori di quanti ne commetterebbero gli uomini medi facendo la stessa attività, allora i residui errori inevitabili di quel sistema dovrebbero essere tollerati senza sanzione penale.

L'autore propone una "presunzione di affidabilità algoritmica" simile alla *business judgment rule* societaria o alla disciplina penal-medica sulle linee-guida: ad esempio, nessuna responsabilità per produttori o utilizzatori di veicoli autonomi che abbiano un tasso di incidenti molto inferiore a quello medio umano (es. <20% del tasso umano).

L'idea è che 'l*a perfezione non è richiesta ove la tecnologia superi l'essere umano in efficienza e sicurezza'*. Questa proposta, per quanto "originale", mira a incentivare l'adozione di tecnologie salvifiche senza la remora del timore di punizioni per gli incidenti sporadici non eliminabili.

Naturalmente sarebbero esclusi dalla franchigia i casi di colpa vera e propria (negligenza nella manutenzione, ignoranza di difetti noti, manomissioni dolose). Il vantaggio prospettato sarebbe così duplice: evitare processi penali nei casi in cui la colpa umana è in realtà minima o assente perché l'algoritmo ha agito entro margini di errore ragionevolmente accettabili, e (al contempo) stimolare lo sviluppo e l'uso di IA che riduca i danni complessivi. La proposta ha ovviamente acceso il dibattito: da un lato, c'è chi la ritiene incompatibile col principio di uguaglianza (creerebbe una sorta di immunità per chi usa IA avanzate, tipicamente grandi aziende) e col diritto delle vittime a giustizia; dall'altro lato, c'è chi intravede in essa una visione pragmatica che riconosce la natura probabilistica del rischio tecnologico. Per ora, si tratta solo di una suggestione dottrinale, ma resta indicativa delle soluzioni non convenzionali pure allo studio[93].

Riforma del nesso causale e prova scientifica

Altra proposta riguarda l'adattamento delle regole sull'accertamento causale ai contesti di IA. Alcuni autori[94] suggeriscono che, nei casi di malfunzionamenti algoritmici, si potrebbe ammettere una prova causale per alta probabilità statistica in luogo dell'esigibilità della certezza *next to certainty* oggi richiesta dal diritto penale. Questo per evitare che l'opacità dell'algoritmo (assenza

[92] F. Lamesso, "*Guida autonoma e responsabilità penale: serve rischiare*", 15/06/2025 <https://www.lexit.it/guida-autonoma-e-responsabilita-penale-serve-rischiare/>
[93] I*vi*
[94] R. Borsari, Intelligenza Artificiale e responsabilità penale: prime considerazioni

di spiegabilità completa) impedisca di dimostrare sempre il nesso oltre ogni ragionevole dubbio. Ad esempio, se il 99% di incidenti come quello in esame è dovuto a un *bug* nel *software*, si potrebbe inferire il nesso anche senza certezza assoluta.

Siffatta linea urta tuttavia con le garanzie dell'imputato e con il rischio di condanne ingiuste basate su mere statistiche. Probabilmente tale proposta non avrà seguito, ma è (anch'essa) sintomatica della tensione tra esigenza di adattare il processo penale alla complessità scientifica e salvaguardia degli standard probatori tradizionali.

In conclusione, le proposte *de iure condendo italiane* si muovono su un crinale sottile: innovare il diritto penale per gestire il rischio dell'IA senza stravolgerne i fondamenti.

Il consenso pare convergere su alcuni punti: mantenere fermo il principio di personalità della responsabilità; potenziare la tutela anticipata con reati di pericolo e obblighi di sicurezza[95]; prevedere meccanismi normativi di esonero o attenuazione della responsabilità quando la tecnologia sia usata correttamente e offra benefici netti; garantire sempre un significativo controllo umano nelle decisioni critiche per non delegare mai integralmente all'algoritmo il destino delle persone.

La dottrina e il legislatore sono così chiamati a un difficile esercizio di equilibrio: l'IA rappresenta una sfida epocale per il diritto penale, ma anche l'occasione per ribadire la centralità dei valori costituzionali (personalità, colpevolezza, legalità) nell'era digitale[96]. E se questo già non bastasse, ogni vano tentativo di strizzare l'occhio all'artificio macchinoso e robotico, seguiterebbe ad avere come termine di confronto: capacità creativa, immaginazione ed emotività, tensione alla critica ed alla risoluzione del dubbio, pensare problematico: che appaiono (tutte) caratteristiche fondanti dell'agire umano. Requisiti non surrogabili da un qualche *software* intelligente anche e soprattutto quando si discuta di responsabilità e pene.

Come ben sintetizzato da un alto magistrato, *«la sfida è coniugare l'evoluzione tecnologica… con le garanzie»*, con la massima attenzione perché *«sono in gioco i diritti fondamentali[97]»*.

[95] *Ibidem*
[96] F. Basile & B. Fragasso, *Intelligenza artificiale e diritto penale: prove tecniche di convivenza*, in Dirittodidifesa.eu, 2023 – sul principio "*machina delinquere non potest*" e antropocentrismo del diritto penale
[97] Giovanni Canzio, primo presidente emerito della Corte Suprema di Cassazione, in occasione di un convegno sulle problematiche della i. a. sulla giustizia, 28 ottobre 2023