

VI sessione - venerdì 2 dicembre 2022, ore 11

**Machina delinquere (et puniri) non potest?**  
*Ambiti, soluzioni e sviluppi del rapporto tra diritto penale e intelligenza artificiale*

(Bozza provvisoria)

**di Cristiano Cupelli**

*Professore ordinario di diritto penale  
Università di Roma Tor Vergata*

1. Come è emerso in maniera nitida nel dibattito di questi giorni, l'intelligenza artificiale (nel prosieguo, anche "IA") permea settori sempre più estesi della nostra quotidianità, favorendo l'inevitabile insorgere di poliedriche questioni giuridiche, che finiscono per coinvolgere anche il diritto penale: quella branca del diritto cioè storicamente più refrattaria all'apertura ai nuovi saperi e, in particolare, alle innovazioni tecnologiche e che, peraltro, come è stato messo in luce, "dinanzi all'IA rischia di apparire un arnese vecchio, ritagliato da sempre sull'uomo in carne ed ossa" (C. PIERGALLINI, 2020, 1746).

La **premessa** rende ancora più meritoria l'occasione di incontro odierna, che rappresenta una testimonianza emblematica di come il rapporto tra IA e scienza penale sia destinato a una vorticoso fase di sviluppo, non foss'altro che per colmare il ritardo accumulato, tanto sul piano della riflessione teorica, quanto su quello legislativo (G. CANZIO, 2021, 3). Volendo, una **conferma** di questo fermento la si può trarre, sul versante normativo, da due recenti provvedimenti: la Risoluzione del Parlamento europeo del 6 ottobre 2021 dal titolo "*L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*", che, pure non avendo forza vincolante, rappresenta un significativo punto di partenza per cogliere l'orientamento dell'Unione europea in rapporto ai profili di responsabilità penale collegati all'IA, da un lato; la proposta di Regolamento europeo (c.d. *IA Act*), che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione COM/2021/206, 21 aprile 2021, focalizzato sui sistemi di IA "ad alto rischio" e sull'individuazione degli obblighi e delle responsabilità che dovrebbero essere poste in capo ai fornitori dei sistemi di IA, dall'altro.

2. Molteplici sono gli ambiti all'interno dei quali la rivoluzione tecnologica messa in moto dall'IA può significativamente impattare sulle pretese di tutela dei beni giuridici affidate al diritto penale (F. BASILE, 2019, 68; M. PAPA, 2019, 3).

In uno sforzo di massima schematizzazione, possono tali ambiti essere compendati in **quattro macroaree**: a) **le attività di law enforcement** (macchine robotiche, utilizzate per una

varietà di compiti presso le forze di polizia di alcuni Paesi, per lo più in via sperimentale, come ad esempio attività di pattugliamento, sorveglianza, disinnescamento di bombe, individuazione di atteggiamenti sospetti, riconoscimento facciale, ecc.) e di **polizia predittiva** (rivolte allo studio e all'applicazione di metodi statistici con l'obiettivo di "predire" *chi* potrà commettere un reato, o *dove* e *quando* potrà essere commesso un reato, al fine di prevenire la commissione di fatti illeciti); **b) il possibile impiego di algoritmi decisionali**, dopo una sperimentazione per la decisione di liti civili o di contenziosi amministrativi (M. LUCIANI, 2018, 872), **anche in ambito penale** (sul punto, il **Consiglio d'Europa**, tramite la propria Commissione per l'efficacia della giustizia, ha espresso serie e motivate preoccupazioni nella *Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti*, adottata il 4 dicembre 2018 dalla CEPEJ); **c) la valutazione, in senso ampio, della pericolosità criminale**, ai fini dell'applicazione di una misura di sicurezza, di una misura cautelare o di una misura di prevenzione o anche per concedere la sospensione condizionale di una pena o l'affidamento in prova al servizio sociale, **affidata ai c.d. algoritmi predittivi** (*risk assessment tools*), capaci di attingere e rielaborare quantità enormi di dati così da far emergere relazioni, coincidenze, correlazioni, che consentano di profilare una persona e prevederne i successivi comportamenti, anche di rilevanza penale (A.M. MAUGERI, 2021, 6); **d) il coinvolgimento di un sistema di IA come strumento, autore o vittima** di un reato.

3. Rappresenta, quest'ultimo, uno dei **punti centrali** della riflessione, su cui pertanto occorre focalizzare l'attenzione.

Gli **esempi** sono svariati e tutti noti: si va dai droni capaci di uccidere per le strade urbane (come avvenuto nella città di Dallas nel luglio 2016) alle auto senza conducente coinvolte nella causazione di incidenti a danno di cose o persone (è ciò che accaduto nel marzo 2018 in Arizona), sino ai *software* che eseguono, in collaborazione o addirittura in sostituzione dell'uomo, compiti sempre più sofisticati (come pilotare un grosso aereo), ma che talvolta possono interferire negativamente con la condotta umana, come i recenti disastri aerei dei Boeing 737 Max hanno purtroppo dimostrato.

**3.1.** La casistica stimola un immediato **interrogativo**: qualora, in episodi siffatti, si riscontri un fatto di reato, chi ne risponderà penalmente?

La risposta più intuitiva e per certi versi 'tranquillizzante' è quella che a rispondere sarà chiamato il programmatore del software, il suo produttore o il suo utilizzatore. Questo approccio si basa in realtà sull'assunto, altrettanto immediato, che il sistema di IA, in fondo, non può che rappresentare uno strumento inanimato in mano all'uomo e attraverso il quale il reato viene realizzato, in ragione delle enormi potenzialità asservibili a scopi criminali. Tra le condotte illecite che potrebbero essere agevolate possono annoverarsi, senza pretesa di esaustività, i crimini informatici, economici ed ambientali, i traffici internazionali di sostanze stupefacenti e di altri prodotti illeciti, la tratta di esseri umani, le violazioni in materia di *privacy*, trattamento dei dati personali, proprietà intellettuale e industriale, oltre alle condotte di diffamazione e di abuso della credulità popolare (magari realizzate attraverso bot che creano *fakenews* destinate alla rete) e alle ipotesi di manipolazione abusiva del mercato, realizzabili attraverso sofisticati programmi informatici, a cui sia riservata non solo l'esecuzione delle transazioni finanziarie, ma anche la stessa decisione di

compierle sulla scorta di un algoritmo idoneo a comparare, in una frazione di secondo, numerose variabili.

**3.2.** Ora, se come si è detto da un certo punto di vista l'attribuzione di una qualche forma di responsabilità a un essere umano appare 'tranquillizzante' rispetto alle nostre 'abitudini' dogmatiche, non possono tacersi talune **significative criticità**; si pensi alla difficoltà di ricostruire la posizione di garanzia di un "sorvegliante umano" (il soggetto cioè investito dall'obbligo e titolare di un effettivo potere di impedimento), alle problematiche sollevate in termini di accertamento del nesso causale (fra l'altro per la possibile presenza di svariati fattori causali alternativi, umani e non) e alle ricadute sul profilo dell'impedibilità dell'evento. Sul versante dell'imputazione colposa, poi, chi parla di colpa da programmazione pone quesiti di non agevole soluzione: come dovrebbe essere configurato questo tipo di colpa? Più nel dettaglio, come dovrebbero essere caratterizzate le regole cautelari di condotta nella realizzazione e nell'implementazione di un sistema di IA? Quale la condotta dell'agente umano modello? E soprattutto, come enucleare un agente umano modello? (A. GIANNINI, 2021, 23).

La difficoltà spinge dunque da un lato a dovere 'riprogrammare' le tradizionali categorie e dall'altro a immaginare nuove fattispecie di reato (o a rimodellare quelle già esistenti) per renderle aderenti alle nuove esigenze politico-criminali e applicabili alle descritte condotte.

**4.** Vi è tuttavia un'ipotesi più complessa e ardita, che chiama in causa questioni più generali, quasi filosofiche: può un sistema di IA assumere la veste di *autore del reato*? E quali sono i **margini di una possibile responsabilizzazione della macchina**?

Si tratta di una prospettiva che si fonda sul presupposto che a essere coinvolto nella commissione di un fatto di reato sia un sistema di IA di ultima generazione, dotato cioè di una qualche capacità di apprendimento e di autonomia decisionale; il che porta a domandarsi se non si possa individuare direttamente nel sistema l'autore del reato, con l'effetto, in caso di risposta affermativa, che dovrebbe essere quest'ultimo a rispondere penalmente.

Orbene, è chiaro che in tutti i casi in cui la condotta dell'uomo si intreccia e si interseca con l'attività di un sistema di IA, il percorso di attribuzione delle responsabilità senza dubbio si complica, giacché le scelte, le valutazioni, i bilanciamenti, sottesi alla commissione del fatto, non sono più opera esclusiva dell'uomo, ma sono quantomeno condivisi con (se non interamente delegati alla) macchina; e se il sistema presenta – come nell'ipotesi in considerazione – autonomia decisionale e capacità di apprendimento e di reazione per effetto della propria esperienza e interazione con l'ambiente, diventa difficile escluderlo dal meccanismo imputativo di responsabilità.

**4.1.** Al di là del piano logico, occorre tuttavia fare i conti con la compatibilità di un'allocazione di responsabilità penali a entità diverse dall'uomo con le peculiari caratteristiche del diritto penale.

Come si è accennato, non manca chi, sul punto, offre una risposta affermativa, facendo leva proprio sul fatto che i sistemi di IA, perlomeno quelli più evoluti e sofisticati, sono capaci di agire in autonomia, di assumere ed eventualmente attuare decisioni proprie, che non erano prevedibili dai loro programmatori (R. BORSARI, 2019, 266). Ci si riferisce, in

particolare, alle tesi avanzate da una parte della letteratura straniera di *common law*, e soprattutto da **Gabriel Hallevy**, incline a ravvisare l'assenza – già oggi – di argomenti ostativi alla perseguibilità e punibilità di soggetti artificiali intelligenti, quali *robot* chirurgici, droni autonomi o *self-driving cars*.

Coerentemente, i fautori di questa impostazione giungono a ritenere che il margine di decisione autonoma di cui dispongono i sistemi possa rappresentare la breccia attraverso la quale potrà farsi strada, in un futuro non lontano, una nuova teoria della responsabilità (anche penale), i cui presupposti e limiti non necessariamente dovranno (e potranno) coincidere con quelli tipici degli esseri umani (in un ardito parallelismo con la disciplina della responsabilità degli enti e con la figura della c.d. colpa di organizzazione, si ipotizzano così i contorni di una "*colpa di programmazione*").

**4.2.** Per quanto suggestiva, questa lettura **non appare, almeno allo stato, del tutto convincente.**

Non si può fare a meno di ricordare, intanto, come il diritto penale, più di tutti gli altri settori dell'ordinamento, sia plasmato sulla fisionomia dell'essere umano: solo gli uomini possono davvero delinquere, scegliendo un male per la società o il prossimo, pur di conseguire un proprio tornaconto personale, e di conseguenza possono ragionevolmente essere puniti in segno di biasimo per la loro scelta colpevole. Inoltre, sul piano passivo, solo esseri umani possono dirsi titolari di beni e interessi giuridici di tale pregnanza da legittimare il presidio di una sanzione idonea a incidere sulla libertà personale dell'individuo. In proposito, parafrasando la classica formula che negava lo *status* di possibile agente delittuoso alle persone giuridiche, è stato persino coniato un brocardo, che rende l'idea di iniziale e radicale chiusura alla configurabilità di soggetti robotici quali attori criminali: *machina delinquere (et puniri) non potest* (A. CAPPELLINI, 2018, 4).

Sul punto, si è osservato come, in fondo, a partire dal 2001, questo **pregiudizio 'ideologico'** sia stato **superato**, venendosi a configurare a tutti gli effetti, con il d.lgs. n. 231 del 2001, una responsabilità da reato in capo agli enti, a carico, quindi, di "persone" che sono tali solo per effetto di una *fictio* giuridica (per l'appunto, "persone giuridiche"); si tratta di una obiezione alla quale, condivisibilmente, si è tuttavia risposto ritenendo il parallelismo improprio o addirittura "malfermo", sulla base di una valorizzazione della "fisicità" posseduta, a differenza delle società, dai sistemi di IA (C. PIERGALLINI, 2020, 1766).

**4.3.** Restando sulle ragioni critiche, ci si può in questa sede limitare a ricordare come la principale sia legata proprio alla difficoltà di ravvisare in capo a una macchina - a differenza degli altri requisiti del reato, quali la commissione di un fatto storico, corrispondente alla previsione di una norma astratta incriminatrice, l'antigiuridicità e la punibilità - il fondamentale requisito della **colpevolezza**. Si tratta, come è ormai acquisito, di un requisito che esprime nella maniera più intensa il coinvolgimento soggettivo dell'autore al fatto commesso e la sua presenza comporta la possibilità di muovere un rimprovero che è prerogativa tipica ed esclusiva di un essere umano; una rimproverabilità che, peraltro, presuppone la ravvisabilità, in capo al soggetto, di imputabilità, dolo o colpa, conoscenza (o per lo meno conoscibilità) della legge penale violata e assenza di cause di esclusione della colpevolezza.

Ebbene: questi elementi, originariamente concepiti e tradizionalmente riferiti solo all'umano, si possono riferire anche a una macchina? Detto altrimenti, si può ipotizzare una

*colpevolezza robotica*, in assenza, fra l'altro, di requisiti *certi* quali l'autocoscienza, il libero arbitrio, l'autonomia morale? Ancora: si può parlare di "capacità di intendere e di volere" in relazione a un software? O configurare una "colpa" (quale mancata osservanza di una regola precauzionale di comportamento) o addirittura un "dolo" (quale volontà consapevole di realizzazione del fatto) in capo ad un sistema di IA?

Proseguendo nel cammino delle criticità: **quali sanzioni potranno essere comminate a tali sistemi?**

Si tratta, a ben vedere, di altra questione cruciale, dal momento che il rilievo penale della responsabilità è ampiamente, pur se non esclusivamente, connesso alla tipologia di sanzioni che possono essere applicate all'esito dell'accertamento della responsabilità: insomma, è possibile parlare di responsabilità penale (solo) se la sanzione applicata è una pena, la quale - nel nostro sistema costituzionale - assume una precisa fisionomia in termini di afflittività e soprattutto riveste una peculiare funzione. Tradizionalmente - è quasi superfluo ricordarlo - si riconosce che una pena possa avere, esclusivamente o cumulativamente e al di là di quella **retributiva** (inflitta cioè per retribuire il male recato con il reato), una **funzione general-preventiva** (la minaccia della pena funge, per i suoi possibili destinatari, da freno alla spinta criminale) o una **special-preventiva** (l'inflizione, e l'esecuzione, della pena pone il condannato nella condizione di non commettere più, in futuro, reato).

Ebbene, di queste funzioni, non manca chi si spinga a ritenere che talune possano essere realizzate nei confronti dei sistemi di IA: attraverso alcune tipologie di 'pene robotiche', quali lo spegnimento definitivo o temporaneo della macchina o la sottoposizione della macchina, dotata di congegni di autoapprendimento, a un nuovo training 'rieducativo', si potrebbero infatti realizzare, rispettivamente, la funzione retributiva e la funzione special-preventiva della pena. Anche su questo profilo, a dire il vero, si addensa più di una nube. Vi è così chi mette in evidenza, da un lato, l'impossibilità di profilare il momento retributivo, giacché questo postula la possibilità, comunque, di muovere un rimprovero 'colpevole', e, dall'altro, l'ancor più ardua realizzazione di qualsivoglia finalità di prevenzione speciale, in ragione dell'assoluta carenza di capacità di autoriflessione da parte della macchina, con preclusione di ogni possibilità di comprendere, attraverso la sanzione, l'errore commesso per evitare di ripeterlo (C. PIERGALLINI, 2020, 1769). Quasi impossibile immaginare, in ogni caso, una funzione general-preventiva, a meno di non concedere spazio a ipotesi per ora ai confini della fantascienza, come una formulazione del precetto penale in termini (digitali) tali che possa essere recepito ed elaborato dalla "comunità" dei sistemi di IA oppure una trasmissione, tramite i canali dell'Internet delle Cose, dell'esperienza della pena subita dal singolo computer, punito in quanto riconosciuto responsabile del reato, anche ai computer consimili, futuri potenziali autori di reati.

5. Alcuni degli argomenti spesi per considerare il sistema di IA quale "persona" (e, quindi, quale autore di reato), potrebbero a ben vedere offrire elementi a favore di un suo riconoscimento quale *vittima del reato*: non solo, quindi, come "cosa" inanimata che subisce materialmente il reato, ma anche come "persona" o come soggetto equiparato alla persona che soffre del reato subito.

A un così impegnativo riconoscimento - quasi una sorta di umanizzazione della macchina - sembrerebbe, tuttavia, frapporsi la considerazione che i sistemi non hanno, né avranno mai, veri sentimenti. Ciò non esclude, tuttavia, una riflessione ulteriore

sull'opportunità di introdurre nuove e apposite figure di reato (o di modificare figure già esistenti), in modo da rendere punibili attacchi rivolti specificamente ai sistemi di IA, che ad oggi non trovano adeguato e pieno riscontro nelle vigenti disposizioni penali.

La casistica a cui fare riferimento, anche qui, è ampia; possiamo limitarci conclusivamente a richiamare ancora l'ipotesi di robot, che riproducono bambole o animali, di cui si comincia a valorizzare l'utilizzo all'interno di programmi di *doll therapy* o *pet therapy*, rivolti a soggetti autistici, malati di Alzheimer, persone con disabilità mentali, rispetto ai quali il paziente potrebbe sviluppare sentimenti e nutrire emozioni (in ciò consistendo lo scopo della terapia): ebbene, la distruzione di questi robot, il loro maltrattamento, la loro sottoposizione malevola a logoramento integra un semplice fatto di danneggiamento? O, per lo meno in quei casi in cui i robot siano dotati di capacità cognitiva e siano specificamente destinati a instaurare una relazione affettiva con il paziente, non ci si muove già in una zona prossima a quella oggi coperta – se non dal reato di maltrattamenti contro familiari e conviventi (art. 572 c.p.), perlomeno – dal reato di maltrattamento di animali (art. 544-ter c.p.)? Per altro verso, si pensi alle preoccupazioni connesse al fenomeno del c.d. “**stupro robotico**” e in particolare agli atti sessuali con robot aventi le dimensioni e le fattezze di minori: se, da un canto, la criminalizzazione di tali condotte parrebbe un'ulteriore concessione ad una visione moraleggiante e paternalistica del diritto penale, dall'altro non dovrebbe rimanere priva di ogni rilievo la tematica del “consenso” del robot all'atto sessuale, ovviamente un consenso da ricostruire e interpretare in termini ben diversi da quelli in cui concepiamo il consenso delle persone umane. *Quid iuris*, ad esempio, nel caso di compimento di atti sessuali con un androide, progettato originariamente non per tale scopo, ma per svolgere funzioni di *receptionist* all'interno di un albergo o di un grande magazzino? Si tratta di ulteriori temi che, forse con minore urgenza rispetto a quanto sopra prospettato, meritano attenzione da parte del legislatore.

6. Concludendo, tutti gli scenari analizzati sono accomunati dall'assenza di una regolamentazione che prevenga o reprima adeguatamente offese penalmente rilevanti. Le ipotesi e le previsioni che sino a poco tempo fa apparivano futuristiche sono oggi divenute realtà, rendendo ingiustificabile tanto il perdurare di una *colposa* sottovalutazione del tema quanto una *dolosa* impreparazione culturale di fondo. Se il diritto penale vuole continuare ad assolvere il suo compito coesenziale, deve mettersi nelle condizioni di affrontare consapevolmente le sfide, complesse ma ineludibili, della modernità e dello sviluppo tecnologico, ripensando schemi classici (senza stravolgere principi consolidati di garanzia; M. LUCIANI, 2018, 872; V. MANES, 2020, 11) e proponendo modelli nuovi di incriminazione e paradigmi di tutela, aperti alle spinte del futuro; va in sostanza scongiurata, su questo terreno, la tentazione di lasciarsi accompagnare (e inevitabilmente travolgere) dall'onda di scelte compiute altrove, per lo più prive di quel corredo di imprescindibili garanzie, sostanziali e processuali, che rappresentano il bagaglio inalienabile della nostra tradizione costituzionale.