TRAIL Series

PAPER N. 4

a.a. 2019/2020

Trento BioLaw Selected Student Papers

Intelligenza artificiale e diritto penale

COSTANZA ROSA MONGIOÌ, GIOVANNI REN, GIACOMO ZANESCO, BIANCA ZIVIANI

Jean Monnet EUROPEAN CENTRE UNIVERSITY OF TRENTO

Co-funded by the Erasmus+ Programme of the European Union

UNIVERSITÀ DI TRENTO Facoltà di Giurisprudenza

I paper sono stati selezionati a conclusione del corso libero *Diritto e Intelligenza Artificiale* a.a. 2019-2020, organizzato all'interno del Progetto Jean Monnet "TrAIL – Trento Artificial Intelligence Laboratory", coordinato presso l'Università di Trento dai docenti Carlo Casonato e Simone Penasa.

Intelligenza artificiale e diritto penale

*Costanza Rosa Mongioì, Giovanni Ren, Giacomo Zanesco, Bianca Ziviani**

ABSTRACT: Technological evolution set by AI could significantly impact legal assets protection entrusted by the criminal law. This revolution goes, mainly, in four direction: predictive policing, where AI systems could help prevent criminal activities; the crime risk assessment tool that, thanks to enormous processing capacity, could allow to profile the subject in order to predict his behaviour; the use of decision–making algorithms to resolve criminal dispute in a sort of replacement, or at least guidance, of the judge-man; finally, the hypothesis of involvement, as a tool or as an author, of an AI system in the commission of a crime.

KEYWORDS: Artificial intelligence; criminal law; predictive policing; risk assessment tool; decision-making algorithms.

SOMMARIO: 1. Introduzione – 2. Polizia predittiva – 3. Algoritmi predittivi della pericolosità sociale – 4. Giustizia predittiva – 5. IA e responsabilità penale. – 6. Conclusioni.

1. Introduzione

L'Intelligenza Artificiale è sempre più pervasivamente presente nelle nostre vite, pertanto è destinata ad avere implicazioni rilevanti anche e soprattutto nell'ambito giuridico, e chiaramente il diritto penale non fa eccezione. Le maggiori problematiche inerenti a questa branca del diritto derivano dal principio di legalità, che pertanto richiede un intervento preciso ed esplicito del legislatore nella regolazione della materia. Come vedremo, il legislatore, soprattutto quello italiano, finora si è mostrato inerte rispetto a queste tematiche, tant'è che l'unica normativa di riferimento è rinvenibile nell'ambito dell'Unione Europea, dove pure abbonda il cosiddetto *soft law*. È pertanto auspicabile una presa di coscienza da parte dei Governi e dei Parlamenti, alla luce delle rilevanti implicazioni penalistiche di tali novità tecnologiche, tra le quali si segnalano: le attività di c.d. polizia predittiva; i c.d. *automatic decision systems*, che in prospettiva potrebbero addirittura sostituire la figura del giudice-persona fisica; i c.d. algoritmi predittivi della pericolosità sociale di un imputato nonché l'ipotesi di coinvolgimento dell'IA nella commissione di un reato.

2. Polizia predittiva

Per "polizia predittiva" si intende l'insieme delle attività volte 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 dei reati stessi[1]. La predizione si basa sulla rielaborazione probabilistica di una serie di dati, che riguardano sia la commissione di reati (luoghi e tempi maggiormente teatro di reati, condizioni atmosferiche, etc.), sia i soggetti che li commettono (origine etnica, livello di scolarizzazione, etc.). L'impiego di *software* basati sull'IA ha reso possibile l'acquisizione e rielaborazione di un'enorme quantità di dati, il che ha portato a un salto di qualità in queste tecniche[2].

Esistono due tipi di *software* di polizia predittiva: quelli basati sugli *hotspots* e quelli basati sul *crime linking*.

Nel primo caso, si tratta di *software* finalizzati a individuare c.d. zone calde o *hotspots*, cioè luoghi che potrebbero essere scenario di futuri crimini, utilizzando acquisizioni della criminologia ambientale[3]. Esempi di sistemi di questo tipo sono *Risk Terrain Modeling* (RTM), un sistema specifico per il perseguimento di reati di spaccio, e *PredPol*, un *software* sviluppato dalla UCLA e in uso diffusamente già da alcuni anni negli Stati Uniti e nel Regno Unito, finalizzato all'individuazione di *hotspots* in relazione a un numero più elevato di reati.

Nel secondo caso, invece, i sistemi, basandosi sul *crime linking*, seguono le serialità criminali di determinati soggetti (individuati o ancora da individuare) per prevedere dove e quando commetteranno il prossimo reato. Attraverso la raccolta e l'analisi di un grande quantità di dati, i sistemi cercano di "profilare" il possibile autore della serie criminale e prevederne le mosse[4]. Peraltro, *software* di questo tipo hanno un'utilità che va oltre la finalità predittiva: infatti possono essere utilizzati per ricostruire la carriera criminale del soggetto "profilato", in modo da potergli imputare non solo l'ultimo reato commesso, ma anche la serie di reati precedenti. In questo caso, tuttavia, sarà necessario che tutti gli elementi presenti nella banca dati siano stati acquisiti rispettando le indicazioni codicistiche[5].

I sistemi di polizia predittiva pongono alcune perplessità dal punto di vista giuridico[6].

In primo luogo, sistemi di questo tipo potrebbero generare problemi rilevanti per quanto riguarda la tutela della *privacy* (in relazione alla grande quantità di dati raccolti) e soprattutto rispetto al divieto di discriminazione, nella misura in cui fattori di pericolosità siano connessi, ed es., a determinate caratteristiche etniche, religiose o sociali. Proprio in relazione a questo problema, una città statunitense

---

* *Studenti dell'Università di Trento, Facoltà di giurisprudenza.*

[1] F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi d'indagine* in *Diritto penale e uomo*, 10, 2019, p. 10.

[2] F. Basile, *op. cit.*, p. 11.

[3] F. Basile, *op. cit.*, p. 11.

[4] F. Basile, *op. cit.*, p. 12.

[5] C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Diritto Penale Contemporaneo*, 6, 2019, p. 58.

[6] F. Basile, *op. cit.*, p. 13.

che si avvaleva del sistema *PredPol* ha in seguito optato per non continuare l'utilizzo[7]. Infatti, se pure il sistema avrebbe potuto portare a una riduzione effettiva del crimine, le conseguenze nefaste sulla relazione fra polizia e determinati gruppi etnici sono state giudicate un rischio troppo elevato, particolarmente in aree in cui questa relazione è stata già esacerbata da passate discriminazioni. In questo frangente emerge dunque la responsabilità che hanno i programmatori di questi sistemi nel momento in cui fissano i parametri d'azione e i dati che sono da prendere in considerazione.

Un'altra criticità che si registra in relazione all'uso di questi sistemi è che in parte si auto-alimentano con dati prodotti dal loro stesso utilizzo, creando un circolo vizioso. In questo modo, si potrebbe arrivare a creare zone altamente sorvegliate, in cui sono dunque rilevati un numero elevato di crimini, e zone non presidiate che possono diventare zone franche per la commissione di reati. Si rileva, in questo senso, che gli operatori di polizia hanno una grande responsabilità nel modo in cui utilizzano questi sistemi, in quanto essendo la componente umana del "meccanismo", dovrebbero essere in grado di analizzare criticamente i dati forniti dai sistemi.

Vista la crescente espansione dei sistemi di polizia predittiva, e in ragione dei potenzialmente gravi rischi ai diritti (anche costituzionalmente garantiti) dei soggetti coinvolti, si nota con preoccupazione la mancanza di una legislazione di riferimento, sia a livello nazionale, sia a livello internazionale. Infatti, tutti i Paesi dove sono stati finora applicati *software* di polizia predittiva (e l'Italia è uno di questi[8]), le modalità dell'utilizzo dei sistemi e dei risultati da essi raggiunti sono state regolate dalla sola prassi dei vari operatori di polizia, che chiaramente costituisce un dato sensibilmente variabile nel tempo e nello spazio, e potrebbe contribuire a peggiorare le criticità rilevate.

3. Algoritmi predittivi della pericolosità sociale

Gli strumenti di *risk assessment tool* analizzano un numero molto elevato di dati relativi al passato (in tema, ad esempio, di percentuali di soggetti recidivi o che non si sono presentati alle udienze dopo essere stati liberati), grazie ai quali è possibile "predire" quali soggetti possono essere rilasciati, magari dietro il pagamento di una cauzione e quali invece devono restare sottoposti a custodia cautelare in carcere[9]. L'utilizzo di questi *tools* pone problemi in ordine, principalmente, al determinismo giuridico, alla trasparenza e alla tutela della privacy.

---

[7] E. THOMAS, *Why Oakland Police Turned Down Predictive Policing*, in *Vice.com*, 2016, link: https://www.vice.com/en_us/article/ezp8zp/minority-retort-why-oakland-police-turned-down-predictive-policing (ultima consultazione: 4/05/2020).

[8] Ad es. *"X-LAW",* sistema basato sugli *hotspots* in uso alla Questura di Napoli; *Keycrime*, sistema di *crime linking* elaborato dalla Questura di Milano.

[9] M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, p. 3.

La prima problematica si riferisce al fatto che questi dispositivi, a fini statistici, prendono in considerazione fattori strettamente correlati all'etnia, come nel caso dell'americano *Correctional Offender Management Profiling for Alternative Sanction* (COMPAS), che mediamente profila gli afroamericani come "*future criminals*" in misura doppia rispetto ai bianchi, ponendo quindi seri problemi di discriminazione[10], nonché problematiche in ordine alla responsabilità per potenziali lesioni della libertà personale determinato da *bias* di matrice etnico-razziale. La seconda questione riguarda il fatto che tali algoritmi sono normalmente oggetto di segreto industriale[11] quindi non sono compiutamente conoscibili né da parte dei soggetti da essi "giudicati" né da parte degli stessi giudici, con ulteriori problemi inerenti, ad esempio, alla motivazione dei provvedimenti giurisdizionali, con l'annessa questione che il giudice potrebbe essere esonerato da responsabilità per errori giudiziari. L'ultima problematica ha a che fare con la tutela della privacy, evidente nel caso dell'inglese *Harm Assessment Risk Tool* (HART). Quest'ultimo, infatti, prende in considerazione dati che normalmente non sono considerati sensibili, quali il codice di avviamento postale dell'imputato, con la conseguenza che diventano conoscibili informazioni che possono favorire atteggiamenti discriminatori (a causa, ad esempio, della particolare concentrazione di determinati gruppi etnici in determinati quartieri). Altre problematiche possono derivare dalla profilazione degli online data[12]. Chiaramente, l'uso di programmi di giustizia predittiva sarà possibile solo in contesti caratterizzati da una certa continuità giurisprudenziale, cioè in un sistema in cui la probabilità che un singolo giudice si discosti dalla tendenza (c.d. trend) è bassa[13].

Nel panorama giuridico europeo e nazionale attuale non vi sono norme che disciplinino l'uso di questi programmi di giustizia predittiva: al massimo si trovano fonti di soft law, come la "Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia", adottata dalla Commissione per l'efficienza della giustizia, che fissa i principi generali che dovrebbero regolare la materia[14]. Tuttavia, non va dimenticato che, a livello europeo, esistono norme imperative e di rango senz'altro superiore che, così come sono formulate, ostano all'uso di questi algoritmi, perlomeno in assenza di un preciso intervento umano. Il riferimento va, in primo luogo, all'art. 5, par. 3, CEDU, il quale stabilisce che "Ogni persona arrestata o detenuta […] deve essere tradotta al più presto dinanzi a un giudice", lasciando presumere la necessità di un contatto fisico o comunque umano tra soggetto ristretto e magistrato, visti anche i sopracitati problemi di conoscenza dei *tools*[15]. Ciò significa, pertanto, che l'uso di algoritmi predittivi non è di per sé illecito, basta che il giudice-

[10] M. Gialuz, *op. cit.*, p. 5.
[11] F. Basile, *op. cit.*, p. 21.
[12] M. Gialuz, *op. cit.*, p. 11.
[13] C. Castelli, C. Piana, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Questione giustizia*, 4, 2018, p. 153.
[14] M. Gialuz, *op. cit.*, p. 12.
[15] M. Gialuz, *op. cit.*, p. 14.

persona fisica abbia sempre l'ultima parola. Ciò è stato affermato anche dalla Corte Suprema del Wisconsin nel caso Loomis[16], nel quale la Corte, di fronte al ricorso di un imputato che lamentava la predisposizione dell'algoritmo a seguire pregiudizi basati sul genere e sull'origine etnica, ha emanato un *warning*, con il quale ha avvertito i giudici a tenere sempre in considerazione il fatto che l'algoritmo (in questo caso lo strumento Compas) effettua valutazioni su base collettiva e non individuale, comportando quindi il pericolo di sovrastimare il rischio di commissione di reati in capo a talune minoranze etniche[17].

Vi è, quindi, una normativa di base di stampo garantista, al fine di evitare abusi la cui sanzionabilità effettiva è tutt'altro che pacifica. A livello di Unione Europea, inoltre, è ostativo all'adozione di queste novità giurisdizionali anche l'art. 22 del Regolamento sulla Protezione Generale dei Dati, c.d. GDPR (2016/679/UE), che, in tema di tutela della riservatezza, dichiara illegittime tutte quelle decisioni basate unicamente su trattamenti automatizzati che producano effetti giuridici negativi sui diretti interessati, laddove per trattamento automatizzato s'intende quell'insieme di azioni che si sviluppano senza alcun coinvolgimento umano nell'arco del processo decisionale[18]. Per quanto riguarda il diritto italiano, ostativi sono sia il principio della presunzione d'innocenza sia la tradizionale sfiducia del legislatore verso le scienze psicologiche e criminologiche. Infatti, nel primo caso si trova difficile ammettere giudizi predittivi di pericolosità contro un presunto innocente, anche se è la legge stessa a prescrivere la necessità di valutare il carattere o la personalità del reo, desunta da atti concreti o dai suoi precedenti penali (art. 274, co. 1, lett. c. c.p.p.), nonché delle sue condizioni di vita individuale, familiare e sociale (art. 133, co. 2 c.p.). Nel secondo caso si fa riferimento al divieto di perizia criminologica di cui all'art. 220, co. 2, c.p.p., derivante dallo scetticismo legislativo verso chi ambisce a studiare il foro interno del reo, essendo le informazioni ricavate da tali scienze difficilmente verificabili[19]. Tuttavia, è dubbia la qualificazione dell'uso dei *tools* come perizia criminologica o psicologica, il cui divieto è già stato criticato dalla sentenza 124/1970 della Corte Costituzionale[20]. Pertanto, in conclusione, è possibile che in Italia vengano usati questi strumenti, purché si tenga conto del fatto che essi sono pur sempre strumenti sviluppati, nella maggior parte dei casi, in sistemi giuridici di *common law* e che il loro utilizzo, non può tenere in considerazione le differenze esistenti rispetto ai sistemi di *civil* law.

4. Giustizia predittiva

L'utilizzo delle nuove tecnologie all'interno del processo penale non è una novità: basti pensare alle banche dati a cui i giudici possono attingere o ai nuovi processi telematici.

[16] Loomis v. Wisconsin, 881, N. W.2d 749 (2016).
[17] F. BASILE, *op. cit.*, p. 21.
[18] M. GIALUZ, *op. cit.*, p. 16.
[19] M. GIALUZ, *op. cit.*, p. 20.
[20] M. GIALUZ, *op. cit.*, p. 22.

In questa sede si intende analizzare un diverso aspetto, non ancora divenuto realtà, ma lungi dall'essere un problema di rilevanza meramente teorica: l'introduzione di algoritmi predittivi nei procedimenti giurisdizionali penali. In dottrina si parla di giudice-robot e giustizia predittiva per indicare la possibilità di prevedere la probabile sentenza, relativa ad uno specifico caso, attraverso l'ausilio di algoritmi[21].

Se questo può portare l'enorme vantaggio di una giustizia rapida, oggettiva ed accessibile, dall'altro lato non si dovrà dimenticare la rilevanza assegnata al canone di ragionevolezza e al bilanciamento tra interessi contrapposti, che permettono di non perdere la specificità del caso concreto.

Ci si deve quindi chiedere quanto l'introduzione di tali sistemi possa essere compatibile con il nostro sistema di valori (costituzionali, europei ed internazionali), per evitare che vi sia una deriva verso un processo di tipo prescrittivo, dettato dal mero lavoro dell'algoritmo. Si corre infatti il rischio di sostituire la decisione umana con una decisione dettata dalla macchina.

Ciò comporterebbe una serie di problematiche, che impongono ai legislatori un'attenta analisi volta ad una regolazione attraverso linee guida chiare, con al centro il rispetto del principio della dignità dell'uomo.

In primo luogo, si pone poi il problema della qualità dei dati e dei possibili *bias*[22] presenti nell'algoritmo, che rischiano di portare a risultati discriminatori. Si pone poi il problema del ruolo dell'informatico e del programmatore: ad una attenta analisi, potrà risultare che saranno coloro che decideranno effettivamente il caso, proprio perché, costruendo l'algoritmo, decideranno i criteri di valutazione del caso stesso. In questo modo verrebbe meno il principio di indipendenza del giudice, che risulterebbe condizionato dalla soluzione presa dall'algoritmo e che potrebbe (per facilità o per suggestione) adattarsi ad essa. In un sistema di *civil law* come il nostro ciò apre uno scenario in cui il precedente assumerebbe valore vincolante, in quanto le decisioni prese sulla base dell'algoritmo presentano sempre soluzioni di continuità fondate sull'analisi della giurisprudenza e prassi precedente, elidendo fortemente la possibilità di *overruling,* finora mai messa in dubbio e prodotta dalla possibile discontinuità dell'interpretazione giurisprudenziale.

L'ultimo punto di criticità attiene all'acquisizione e alla valutazione delle prove: la non conoscibilità dell'algoritmo (cd *black box*) e la conseguente inaccessibilità della prova ne rendono opaca la genesi in violazione del principio del giusto processo, essendo le decisioni non contestabili, a discapito tanto del principio del contraddittorio (che vede venire meno il principio della parità delle armi a favore dell'accusa e

[21]A. Traversi, *Intelligenza artificiale applicata alla giustizia: ci sarà un giudice robot?* in *Questione giustizia*, 10 aprile 2019, link: http://www.questionegiustizia.it/articolo/intelligenza-artificiale-applicata-alla-giustizia-ci-sara-un-giudice-robot-_10-04-2019.php (ultima consultazione: 4/05/2020).

[22]Sui numerosi problemi concernenti i bias si veda: FLUEL, *Il bias e l'IA*, in *Intelligenza Artificiale Italia (AI)*, 12 ottobre 2017, link: http://aiitalia.it/2017/10/12/il-bias-e-lai/ (ultima consultazione: 12/02/2020).

con il rischio che si perda l'interlocuzione personale delle parti davanti ad un giudice terzo ed imparziale) quando del diritto di difesa.

In quest'ottica, bisognerà lavorare proprio sul fattore "trasparenza dell'algoritmo" e sulla sua conoscibilità non solo da parte di entrambe le parti in causa, ma anche del giudice stesso, per evitare che si addivenga a decisioni che mancano di motivazione e di spiegazione dell'iter logico dietro di esse, con il rischio conseguente che divengano immediatamente (e giustamente) impugnabili e poste nel nulla.

Emerge quindi una necessità sempre maggiore di regolare queste tecnologie che caratterizzano questa nuova era c.d. digitale[23]. Un tentativo di risposta alle difficoltà che intervengono nel momento in cui l'AI viene utilizzata a supporto dell'attività giudiziaria è stato dato dal Consiglio d'Europa con l'adozione della Carta Etica europea per l'uso dell'intelligenza artificiale nei sistemi giudiziari e nei loro ambienti, redatta dalla European Commission for the Efficiency of Justice (2018)[24]. Essa individua *c*inque principi fondamentali che devono guidare gli utenti in senso ampio nell'utilizzo dei meccanismi di IA[25], primo fra tutti il rispetto dei diritti fondamentali, sin dal momento iniziale di design dell'algoritmo, conferendo tutela al principio di non discriminazione. Viene poi posto l'accento sul principio di qualità e sicurezza dei dati, dovendo le decisioni giudiziarie provenire da fonti (dati) certificate e conservate in un ambiente tecnologico sicuro. Segue il principio di trasparenza, imparzialità e correttezza, che andrà correttamente bilanciato con il diritto di proprietà intellettuale. Ad oggi sono in atto proposte di predisposizione di Autorità pubbliche indipendenti che garantiscano la bontà e la qualità dei dati e che possano ispezionare tali si sistemi[26]. L'ultimo principio è il cd *"under user control"*, che vuole quindi il giudice quale *deus ex machina* della decisione, al quale il sistema di AI funge da mero ausilio.

---

[23] Come ricorda L. VIOLANTE: **«**Nel nostro mondo convivono due società molto diverse tra loro, la società analogica alla quale apparteniamo noi e quella digitale alla quale appartiene la generazione sotto i 30 anni. Società che hanno dati differenziali profondi. Nella società digitale si azzera il tempo e lo spazio. La società analogica invece vive nel tempo e nello spazio. Per ragioni anagrafiche la società analogica è destinata ad estinguersi. Quella digitale sarà invece la società del futuro. Abbiamo allora il problema di occuparci di quali saranno le regole in questa società, problema tutt'altro che secondario. C'è certamente il confronto tra conservazione e innovazione. Ma bisogna conservare i valori ed innovare le tecniche» in *Intelligenza artificiale, etica e regole: intervista al presidente Luciano Violante*, in *Fondazione Leonardo - Civiltà delle Macchine,* 9 dicembre 2019, link: https://fondazioneleonardo-cdm.com/it/news/intelligenza-artificiale-etica-e-regole-intervista-al-presidente-luciano-violante/ (ultima consultazione 4/05/2020).

[24] EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment,* 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018).

[25] Per un approfondimento sul tema si veda S. QUATTROCCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un urgente discussione tra scienze penali ed informatiche* in *Legislazione Penale*, 18 dicembre 2018, link: http://www.lalegislazionepenale.eu/wp-content/uploads/2019/02/Carta-etica-LP-impaginato.pdf (ultima consultazione: 4/05/2020).

[26] Per una proposta in tal senso si veda FONDAZIONE LEONARDO - CIVILTÀ DELLE MACCHINE, *Statuto etico e giuridico dell'IA,* 2019, p. 68 ss.: «La collocazione ideale di un'autorità di questo tipo è sovranazionale, perché di rilevanza internazionale sono i temi, le tecniche, i caratteri delle questioni affrontate. Ma la stessa è suscettibile di essere collocata anche a livello domestico, pur sempre entro un network europeo sul modello di altre regolazioni economiche. A una simile Autorità potrebbero essere attribuiti poteri sia amministrativi che "quasi legislativi" e "quasi giudiziali", ricorrendo a tecniche di hard e soft law, e cumulando diverse funzioni che allo stato non appaiono univocamente ascrivibili agli attori pubblici esistenti.».

Il problema principale di questo documento, e di altri simili, è dato dalla sua assenza di vincolatività, essendo un mero strumento di *soft law* che individua certamente delle linee guida condivisibili, le quali però necessiteranno, e in tempi auspicabilmente brevi, di una attuazione di dettaglio che permetta di mantenere salvi quei principi (processuali e valoriali) che caratterizzano tutte le società democratiche.

5. IA e responsabilità penale

Software dotati di IA stanno penetrando sempre più nelle nostre vite; basti pensare alle vetture a guida autonoma, agli algoritmi che eseguono compiti estremamente sofisticati come pilotare un aereo e ai droni che iniziano a circolare in alcune città. Entrando in contatto con l'uomo, direttamente o indirettamente, questi software possono arrecargli un danno e talvolta, integrare una fattispecie di reato. Possiamo, in questi casi pensare al sistema di IA come ad un mero strumento inanimato per la realizzazione del reato o può essere considerato responsabile penalmente? Nel 1981 un impiegato in un'azienda giapponese è stato ucciso da un robot che stava lavorando di fianco a lui; il robot l'ha erroneamente identificato come una minaccia al suo lavoro e l'ha spinto verso un macchinario in funzione lì vicino, uccidendolo[27]. Chi può essere ritenuto responsabile di questo omicidio? Può esserlo direttamente la macchina?

Il più fervido sostenitore di una responsabilità penale delle macchine è Gabriel Hallevy che ha proposto tre differenti modelli di imputazione di responsabilità presupponendo una personalità giuridica degli agenti di IA[28]. Questi modelli comportano una diversa allocazione di responsabilità tra i soggetti coinvolti. Nel primo modello, *perpetration through another*, le IA possono essere assimilate ad un incapace naturale o ad un bambino, difettando quindi di un elemento soggettivo per l'attribuzione di responsabilità. Possono essere però usate come tramite per una condotta illecita da parte di un soggetto pienamente imputabile: il programmatore o l'utilizzatore finale. Non c'è dubbio che in questo modello la responsabilità sia attribuibile al soggetto umano. Il secondo modello, *natural probable consequence*, si basa sulla relazione tra le attività eseguite dalla IA e dall'utente, i quali, non hanno intenzione di commettere reato. Reato che non è stato programmato o cercato dal soggetto umano. Il modello in parola trova quindi applicazione in tutti quei casi in cui è possibile muovere un rimprovero all'uomo per non avere previsto come probabile e naturale il reato commesso. L'elemento psicologico umano è quello della colpa. Tuttavia, possono occorrere due scenari: nel primo, il programmatore/utente è negligente nella programmazione ma senza volontà di reato; la responsabilità viene quindi addebitata per pura colpa per non aver adottato tutte le cautele necessarie.

[27] R. WHYMANT, *Robot kills factory worker*, in *The Guardian*, 9 Dicembre 1981, link: https://www.theguardian.com/theguardian/2014/dec/09/robot-kills-factory-worker (ultima consultazione: 4/05/2020).

[28] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal,* 4, 2, 2010, p. 171.

Nel secondo invece il programmatore/utente ha la volontà di commettere un reato e utilizza quindi dolosamente la macchina per il suo scopo illecito, ma il decorso causale devia, portando l'IA a commettere un reato diverso o un altro reato in aggiunta a quello voluto, integrando quindi lo schema dell'*aberratio delicti*, monolesiva o plurilesiva, portando all'assunzione di responsabilità a titolo di colpa, se il fatto è previsto dalla legge come delitto colposo nel caso di *aberratio delicti* monolesiva, ovvero secondo le regole previste per il concorso di reati, nel caso di *aberratio delicti* plurilesiva; responsabilità che ricadrà in capo all'agente umano[29]. Nel modello *natural probable consequence*, l'IA può "comportarsi" in due modi diversi: come soggetto innocente, andando quindi ad integrare il modello della *perpetration through another*; come parte attiva della condotta criminosa, aggiungendo quindi la propria responsabilità a quella dell'agente umano.

Il terzo paradigma è quello della *direct liability* dell'IA. È possibile integrare, da parte della macchina, la condotta illecita e l'elemento soggettivo? Per quanto riguarda la condotta, nessun problema si pone in quanto la macchina dotata di IA può tranquillamente porre in essere una condotta criminale, sia commissiva che omissiva (si pensi al movimento di un braccio robotico o all'inerzia in una determinata circostanza). Più complessa è l'attribuzione dell'elemento soggettivo. L'IA è dotata, in un certo senso, di "coscienza" in quanto può recepire dati dal mondo esterno, analizzarli ed elaborare una risposta basata su processi cognitivi "intelligenti" assimilabili a quelli umani, necessari per l'attribuzione dell'elemento soggettivo al soggetto agente. In questo senso quindi, integrando sia la condotta che l'elemento soggettivo, le IA possono essere responsabili penalmente delle diverse condotte che pongono in essere[30].

Secondo questi modelli quindi, è configurabile una responsabilità penale dei sistemi di IA che, a seconda dei casi, può essere associata a quella dell'agente umano, ma può anche essere indipendente. Un profilo altrettanto problematico che emerge è quello della pena[31]. Anche volendo accettare queste categorie e delle pene specifiche per le IA, il problema principale riguarda la finalità della pena stessa e con lo scopo che questa persegue negli ordinamenti di *civil law*. Dal punto di vista general-preventivo, appare difficile immaginare la "comunità delle IA" rappresentarsi le conseguenze di condotte riprovevoli commesse dai loro simili. Dal punto di vista special-preventivo poi, è ancora più dubbio di come possa essere rieducata un'IA e imparare quindi a comportarsi conformemente (si pensi alla funzione rieducativa richiesta dal nostro art. 27 Costituzione).

[29] L'art. 83 c.p. comma 1 così recita: «[…] se, per errore nell'uso dei mezzi di esecuzione del reato, o per un'altra causa, si cagiona un evento diverso da quello voluto il colpevole risponde, a titolo di colpa, dell'evento non voluto, quando il fatto è preveduto dalla legge come delitto colposo». "A titolo di colpa" si limita a dire che l'evento non voluto viene punito come se questo fosse colposo comportando quindi, in realtà, una responsabilità oggettiva in capo all'agente.

[30] M. BASSINI, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. PIZZETTI, *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, p. 367.

[31] F. BASILE, *op. cit.*

6. Conclusioni

Dalle quattro analisi effettuate emerge chiaramente la mancanza di una legislazione di riferimento unitamente a una normativa attuale che, così come è interpretata e applicata, risulta ostativa all'applicazione di queste nuove tecnologie nell'ambito processual-penalistico.

Emerge inoltre come una delle chiavi di lettura del rapporto tra diritto penale e IA sia la responsabilità, intesa sia dal punto di vista dell'uomo, responsabile della programmazione degli algoritmi, del loro uso ed eventuale abuso, sia anche (potenzialmente) della macchina stessa, qualora commettesse (più o meno autonomamente) o fosse usata per commettere un reato.

Deve essere in ogni caso chiaro che l'approccio all'IA non può prescindere dalla considerazione che si tratta di una tecnologia fallibile, che non promette risposte o risultati dotati di assolutezza e certezza; alla luce di ciò, è bene che l'intervento e il controllo umano siano sempre valorizzati e garantiti, anche e soprattutto a livello legislativo.