**L’A.I. può subire o commettere reati?**

di Antonio Carlo Oliveri del Castillo

(Estratto dalla tesi di laurea in diritto penale “La responsabilità penale delle intelligenze artificiali”, Luiss, anno 2020-21)

**Premessa. Le sfide future.**

**di Roberto Oliveri del Castillo**

“La mente umana non è, come ChatGPT e i suoi simili, una macchina statistica e golosa di centinaia di terabyte di dati per ottenere la risposta più plausibile ad una conversazione o la più probabile ad una domanda scientifica.(…) La mente umana è un sistema sorprendentemente efficiente ed elegante che opera con una quantità limitata di informazioni. Non cerca una correlazione di dati, ma cerca di creare spiegazioni.” (Noam Chomsky, New York Times, 8 marzo 2023).

La frase di Noam Chomsky è estremamente significativa. Essa fa comprendere meglio di tante definizioni la differenza tra mente umana e I.A., ma dà anche la dimensione e la direzione di una delle sfide più importanti degli ultimi anni e certamente dei prossimi, ovvero quella della natura della intelligenza artificiale e del rapporto che va stabilito tra questa, i rapporti umani e il diritto. Un rapporto che investe i campi più disparati e di conseguenza quello della civiltà giuridica che attiene alle regole dei rapporti umani e sociali. In questo ambito va pertanto esplorato quali sono gli strumenti giuridici per regolare questo nuovo ambito delle attività umane, ovvero quello che vedono la mediazione e l’azione di soggetti non umani, ma dall’uomo impostati e programmati. I campi umani investiti dallo tsunami dell’intelligenza artificiale (ormai per convenzione A.I.) sono i più disparati, dalla salute, alla giustizia, all’istruzione, al lavoro, alla sicurezza, tutti più o meno afferenti la sfera più intima e fondante delle moderne democrazie e dei paesi emergenti. Gli scenari prefigurati possono anche essere, da questo punto di vista, particolarmente inquietanti, fino a mettere in discussione l’essenza stessa dell’identità umana singola e nelle formazioni sociali. Si tratta di capire, in sostanza, come adattare i principi strutturali dei nostri ordinamenti democratici alla novità tecnologica di questi ultimi anni, senza che ciò intacchi le garanzie democratiche e ci faccia sconfinare nelle distopie immaginate nella letteratura nella prima parte del secolo scorso, che infatti avevano come componente fondante una spiccata vocazione tecnologica futuristica distopica, da Metropolis di F.Lang, a Noi di E. Zamjatin, al Mondo Nuovo di A.Huxley ai più noti 1984 di G.Orwell e Il cacciatore di androidi (Blade Runner) di Philip K.Dick, tutti testi in cui l’automazione intelligente è protagonista assoluta.

**Che cosa si intende per Intelligenza artificiale.**

I sistemi di Intelligenza Artificiale rispecchiano il nucleo più profondo della rivoluzione digitale attualmente in atto. Essi rappresentano la frontiera oggi conosciuta, l’apice dell’innovazione tecnologica che l’uomo è stato in grado di conquistare, ma questo non esclude che tale frontiera non diventi invalicabile nel prossimo futuro: come negli anni ’50 e ‘60 le tecnologie moderne potevano esistere solo nell’immaginario collettivo e i primi computer comparsi costituivano il limite allora insormontabile, allo stesso modo le attuali Intelligenze artificiali potranno in futuro raggiungere livelli che adesso sarebbe possibile solo ipotizzare.

L’Intelligenza Artificiale ha fatto la sua prima comparsa nell’estate del 1955, quando uno dei suoi padri fondatori, John McCarthy, assistente universitario di matematica al Dartmouth College di Hanover, organizzò un convegno sul tema dell’Intelligenza Artificiale, avente ad oggetto la dimostrazione che ogni aspetto della mente umana possa essere simulato e replicato da una macchina robotica. L’obiettivo, appunto, era provare che anche le macchine sono in grado di acquisire la capacità – apparentemente appartenente soltanto all’uomo – di ragionare in modo autonomo, di elaborare concetti logici e risolvere problemi.

Più tardi, dopo che furono compiuti notevoli passi in avanti, Roger Schank, uno dei maggiori teorici dell’Intelligenza artificiale, individuò alcuni fattori in presenza dei quali sarebbe stato possibile riconoscere un sistema di Intelligenza Artificiale: la consapevolezza di sé, la consapevolezza della realtà circostante, la capacità di tenere una condotta teleologicamente orientata, volta cioè alla realizzazione di un determinato obiettivo, ed infine la capacità di improvvisare decisioni e iniziative non predeterminate, al mutare della realtà esterna, qualora il piano d’azione prestabilito dovesse fallire o rivelarsi inadeguato. Tali caratteristiche parrebbero sufficienti per sgombrare il campo da alcuni radicati luoghi comuni: in primo luogo, non necessariamente un sistema d’Intelligenza artificiale dovrà identificarsi con un umanoide – e cioè un dispositivo robotico avente sembianze umane e che tuttavia costituisce una delle innumerevoli manifestazioni dell’IA – potendosi trattare anche di semplici algoritmi; in secondo luogo, sembrerebbe più suggestiva che reale un’effettiva somiglianza tra la mente umana e i meccanismi interni che governano il funzionamento di un sistema d’Intelligenza artificiale, ragione per la quale sarebbe preferibile parlare di “razionalità” piuttosto che di “intelligenza”, dovendosi intendere per “razionalità” la capacità di individuare e compiere l’azione migliore in un ventaglio di azioni possibili sulla base dei criteri di economicità ed efficacia.

Ad ogni modo, non esiste oggi una nozione di Intelligenza artificiale che sia univoca e universalmente condivisa. Piuttosto, sono numerose le definizioni che ne sono state offerte da una molteplicità di fonti: << L’Intelligenza artificiale (IA) indica sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull’Intelligenza artificiale possono consistere solo in software che agiscono nel mondo virtuale (ad esempio, assistenti vocali, software per l’analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure incorporare l’Intelligenza artificiale in dispositivi hardware (ad esempio, in robot avanzati, auto a guida autonoma o droni) >> (Comunicazione elaborata dalla Commissione europea, Artificial Intelligence for Europe, 2018.).

Ancora, un’ulteriore nozione è stata elaborata da un team di cinquantadue esperti in materia – definito Gruppo Indipendente e nominato dalla Commissione europea con funzioni di consulenza sull’Intelligenza artificiale – che identifica i sistemi di Intelligenza artificiale in << sistemi software (ed eventualmente hardware) progettati dall’uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l’acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulla conoscenza o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l’obiettivo dato. I sistemi di Intelligenza artificiale possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando gli effetti che le loro azioni precedenti hanno avuto sull’ambiente. Come disciplina scientifica, l’Intelligenza artificiale comprende diversi approcci e diverse tecniche, come l’apprendimento automatico [.], il ragionamento meccanico (che include la pianificazione, la programmazione, la rappresentazione delle conoscenze e il ragionamento, la ricerca e l’ottimizzazione) e la robotica (che comprende il controllo, la percezione, i sensori e gli attuatori e l’integrazione di tutte le altre tecniche nei sistemi ciberfisici >> (Una definizione di IA: principali capacità e discipline scientifiche, elaborato dal Gruppo indipendente di 52 esperti nominato dalla Commissione europea, Bruxelles, 25 aprile 2018).

Altra autorevole dottrina (M. SOMALVICO, ingegnere e accademico italiano, specialista nel campo dell’Intelligenza artificiale) concepisce l’Intelligenza artificiale come quella disciplina, propria dell’informatica, che analizza i fondamenti teorici e pratici che sono alla base dell’elaborazione di sistemi software o hardware in grado di compiere operazioni che tradizionalmente, nell’immaginario comune, solo la mente umana è capace di concepire e porre in essere. Infine, sempre allo scopo di definire in termini chiari un sistema d’Intelligenza artificiale, nell’ottica di fornire delle conoscenze di base utili all’indagine che ci si accinge a svolgere, è possibile richiamare anche la proposta di Regolamento della Commissione europea in materia di Intelligenza artificiale, il cui art. 3 offre appunto una definizione di tale fenomeno. In particolare, un sistema di Intelligenza artificiale viene descritto come un software che, al fine di conseguire uno o più obiettivi che vengono prestabiliti dal programmatore umano, viene progettato con una o più tecniche descritte nella Proposta di Regolamento in questione, e programmato per lanciare output destinati a modificare l’ambiente esterno: in questo senso << Un sistema di Intelligenza artificiale è un software [.] che può, per una determinata serie di obiettivi definiti dall’uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono>> (Art. 3 della Proposta di Regolamento della Commissione europea in materia di Intelligenza artificiale).

In sintesi, sarebbe corretto affermare che per Intelligenza artificiale si intende l’ultima frontiera dell’innovazione tecnologica volta a rendere le macchine indipendenti

**In che modo l’Intelligenza artificiale incide sul diritto.**

Il tema della responsabilità penale delle Intelligenze artificiali per i reati da esse commessi rappresenta oggi una materia senza dubbio ancora non pienamente al centro del dibattito pubblico, ma sicuramente all’avanguardia, trovandosi la società moderna alle porte di una nuova era digitale, destinata a travolgere con tutta la sua forza il mondo oggi conosciuto.

Questo scenario, dunque, impone agli studiosi del Diritto di analizzare a fondo tale fenomeno, di interrogarsi sulle sue implicazioni di ordine giuridico, e di prospettare risposte adeguate alle questioni di rilevanza penale che tale realtà pone, affinché la società possa essere preparata ad affrontarne ogni aspetto quando quello che oggi è presente solo nell’immaginario collettivo o in alcune specifiche realtà più avanzate, diventerà abituale e quotidiano.

L’essere umano, ormai già da diversi decenni, vive sulla propria pelle una nuova rivoluzione industriale avente ad oggetto l’elaborazione di tecnologie sempre più avanzate e sofisticate e che procede a ritmi sempre più rapidi e impetuosi, tanto da aver fatto della tecnologia il nucleo essenziale della propria vita quotidiana, privata e non. Il mondo di oggi è ormai completamente dipendente dall’uso della tecnologia in ogni suo ambito, ed ora, in ultima istanza, si avvia ad esserlo anche in campo giuridico.

<<Nell’arco dei prossimi cento anni, l’intelligenza dei computer supererà quella degli esseri umani>>(Intervento di S. Hawking durante la Conferenza Zeitgeist, Londra, maggio 2015.). Potrebbe sembrare pura retorica, o mera fantascienza, ma l’affermazione del fisico, astrofisico e matematico Stephen Hawking manifesta con estrema semplicità e chiarezza la direzione verso cui inevitabilmente si dirige la società moderna grazie all’inarrestabile progresso tecnologico.

Già da diverso tempo, infatti, la società risulta caratterizzata dal ricorso sempre più frenetico all’utilizzo di sistemi di Intelligenza Artificiale e più in generale dei computer nei più svariati ambiti della vita quotidiana – nella sfera privata, negli ospedali, nelle banche, nelle compagnie di assicurazione – e si avvia ad esserne dominata man mano che vengono elaborate tecnologie sempre più avanzate e raffinate, << e quando questo accadrà, dovremo assicurarci che i computer condividano i nostri stessi obiettivi>> (Intervento di S. Hawking, citato supra, nota 1). Il progresso irrompe senza chiedere il permesso, senza concedere il tempo di elaborare un nuovo modello di società che sia in grado di accogliere con consapevolezza le sue conseguenze, travolgendo gli equilibri su cui attualmente essa si fonda e rivoluzionando i rapporti oggi conosciuti tra uomo e macchina.

L’uomo si ritrova così in balia della tecnologia, e in uno scenario di tal genere, per certi versi allarmante, egli rischia di venirne travolto e di essere del tutto sguarnito dei presidi tradizionali di protezione, concepiti dal legislatore per proteggerlo da condotte umane, e non dalle attuali e future illecite manifestazioni dell’innovazione tecnologica. D’altra parte, la veridicità e lucidità di tali previsioni parrebbe essere confermata anche dall’affermazione contenuta nei Considerando della Risoluzione del Parlamento Europeo sulla Robotica del 16 febbraio 2017 (Risoluzione del Parlamento europeo del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti nome di diritto civile sulla robotica), che dichiara che è possibile che a lungo termine l’Intelligenza Artificiale superi la capacità intellettuale umana.

Proprio in tali considerazioni è possibile scorgere allora una sollecitazione nei confronti del giurista ad indagare a fondo le correlazioni tra Intelligenza Artificiali e Diritto – in particolar modo il Diritto penale — nella misura in cui si afferma che se da sempre la mente umana ha spaziato dando libero sfogo all’immaginazione circa la possibilità di costruire macchine intelligenti, oggi davvero la civiltà umana si trova agli inizi di nuova rivoluzione industriale, di cui sono protagonisti esclusivi robot, bot, androidi ed ogni altra manifestazione di Intelligenza Artificiale, e che è destinata inevitabilmente a coinvolgere ogni aspetto della vita dell’uomo, divenendo quindi imprescindibile che la legislazione ne consideri le implicazioni e le conseguenze legali ed etiche, senza che venga ostacolata l’innovazione. A questo punto, come dinanzi ad ogni progresso nel corso della storia dell’uomo, l’attuale andamento pone un complesso scenario: l’elaborazione e lo sviluppo di macchine robotiche sempre più autonome e intelligenti, capaci di apprendere e adottare decisioni in modo indipendente e sulla base della propria stessa esperienza, comporta non solamente immensi vantaggi in termini economici e di efficienza, ma anche innumerevoli perplessità di carattere giuridico legate agli effetti pratici che l’utilizzo dei sistemi elaborare processi logici prestabiliti dal produttore o di rispondere ai suoi comandi, ma anche capaci di apprendere dalla propria stessa esperienza, di “pensare” e di “comportarsi” autonomamente?

Ad ogni modo pare opportuna, a titolo introduttivo, una sintetica rassegna di quelli che sono i principali ambiti di rilevanza penale sui quali potrebbe incidere con maggior forza la rivoluzione innescata dalla diffusione delle Intelligenze Artificiali.

Tali settori, potenzialmente toccati dal sempre più frenetico ricorso ai sistemi di Intelligenza Artificiale nella moderna società, vengono identificati da autorevole dottrina (F. BASILE, Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine, in Diritto penale e uomo — DPU, 29 settembre 2019), in primo luogo nelle attività di Law Enforcement, ed in particolare modo di polizia predittiva (come non ricordare, sul tema, ad esempio il film di S. Spielberg Minority Report del 2002, tratto dal famoso romanzo di P. K. Dick?), in relazione alle quali le Intelligenze Artificiali sono in grado di offrire un prezioso contributo nel tentativo di contrastare, o anche di prevenire, la commissione di fatti criminosi, mettendo in correlazione una molteplicità di dati provenienti da fonti eterogenee; in secondo luogo nell’affiancamento, se non addirittura nella piena sostituzione, del giudice-macchina al giudice-uomo attraverso l’utilizzo di algoritmi decisionali; ancora, nell’impiego di algoritmi predittivi in grado di individuare e mettere in relazione tra loro enormi quantità di dati personali relativi all’individuo di cui si intende valutare la pericolosità sociale, al fine di far emergere coincidenze e correlazioni che, se osservate attentamente, consentirebbero di disegnare il profilo criminale del soggetto in esame e di ipotizzarne, con un grado di probabilità più o meno elevato, i successivi comportamenti di rilevanza penale; infine – ed è il tema che più interessa ai fini della presente trattazione – nel possibile coinvolgimento dei sistemi di Intelligenza Artificiale in qualità di strumento, di vittima, o di autore nella commissione di un reato.

Per quanto riguarda il primo dei quattro campi d’indagine – l’attività di Law Enforcement ed in particolar modo di polizia predittiva – il ricorso ai sistemi di Intelligenza Artificiale rappresenta già oggi una concreta realtà in atto, di cui si può prospettare una potente crescita e intensificazione nel prossimo futuro: nel documento del 2019 “Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement: An Ally or an Adversary?”, dedicato per l’appunto al tema delle Intelligenze Artificiali e delle attività di Law Enforcement, si dichiara che “Nei loro sforzi per aumentare l’efficienza e l’efficacia e per stare al passo con le innovazioni quadro, tale da evidenziare chiaramente la pericolosità criminale dell’individuo e quindi la necessità di adottare, da parte dell’uomo, concreti provvedimenti volti a prevenire il verificarsi dell’evento terroristico”.

Ad ogni modo, sono diverse le modalità con cui i sistemi di Intelligenza Artificiale possono essere adoperati in questo campo: si tratta principalmente di macchine robotiche, umanoidi e non, utilizzate con un’ampia varietà di funzioni, come attività di sorveglianza, di riconoscimento facciale, di pattugliamento, di individuazione di atteggiamenti sospetti, di disinnesco di bombe, etc.

E tuttavia di tali impieghi dei sistemi di Intelligenza Artificiale è possibile prospettare una serie di controindicazioni che inevitabilmente fanno da contraltare agli immensi benefici: se da un lato è infatti innegabile il merito di preservare da una serie di fonti di pericolo l’agente umano, il quale ha la possibilità di non esporsi in prima linea, nonché il merito di garantire un elevato livello di efficienza e precisione delle prestazioni poste in essere dalla macchina robotica, dall’altro lato non possono trascurarsi una serie di perplessità legate al loro utilizzo.

Un primo punto interrogativo riguarda l’intensità più o meno elevata del controllo da esercitarsi sui sistemi di Intelligenza Artificiale da parte dell’operatore umano, che potrà solo limitarsi all’individuazione degli obiettivi da realizzare, o potrà invece supervisionare più profondamente lo svolgimento della prestazione da parte della macchina robotica, anche a rischio di comprometterla.

In secondo luogo, sorge il rischio di violazione del diritto alla privacy dell’individuo, in considerazione della capacità di tali sistemi algoritmici di acquisire ed elaborare enormi quantità di dati ed informazioni relative alla vita – anche privata – dei cittadini. Infine, bisogna considerare che spesso tali dispositivi risultano equipaggiati con armi letali e non, e da tale circostanza scaturiscono profonde preoccupazioni in merito al grado di affidabilità e fallibilità di tali applicazioni, in merito al rischio di commissione di uccisioni o lesioni per errore, ed in merito all’individuazione della relativa responsabilità in capo alla macchina o all’agente umano.

Rimanendo nell’ambito dell’attività di Law Enforcement, merita un attento e approfondito sguardo il tema della Polizia predittiva, per la quale si intende il complesso di attività volte appunto a predire chi potrà commettere un reato, o dove e quando potrà essere commesso, al fine di prevenirne la verificazione.

Tale previsione si basa principalmente sull’acquisizione, rielaborazione, e messa in relazione tra loro, di diverse tipologie di dati, che possono essere relativi a precedenti penali, ai luoghi, orari e periodi dell’anno maggiormente connessi alla commissione di determinati fatti criminosi, agli spostamenti e alle attività svolte da soggetti sospettati, o ancora dati inerenti al livello di scolarizzazione, all’origine etnica, alle caratteristiche somatiche, o alle condizioni economiche di un individuo. In altre parole, il sistema di polizia predittiva basa il suo funzionamento su dei software in grado di rielaborare enormi quantità di dati e informazioni, scoprendo ed evidenziando correlazioni difficilmente individuabili dal nudo occhio dell’operatore umano. Tali software di polizia predittiva si suddividono in due principali categorie: in primo luogo quelli che si ispirano alla criminologia ambientale e individuano i luoghi dove è più elevata la probabilità che si possa verificare la commissione di determinati reati, vale a dire le cosiddette “zone calde” (“hotspot”); in secondo luogo quelli che si ispirano al concetto di crime linking e monitorano non tanto i luoghi a rischio, quanto i comportamenti di determinati soggetti, al fine di prevedere dove e quando potrebbero commettere il prossimo reato. Ovviamente, tali algoritmi funzionano oggi non in maniera generalizzata, ma solo in relazione alla previsione di specifiche categorie di reati (come rapine o spaccio di stupefacenti, attinenti quindi alla criminalità da strada) – quei reati, cioè, rispetto ai quali assumono particolare rilevanza determinate caratteristiche (come luoghi o fasce orarie in cui è più probabile che vengano commessi), le quali vengono captate e rielaborate dal software ai fini della predizione.

Rientra nella prima categoria di software di polizia predittiva il cosiddetto “Risk Terrain Modeling (RTM)”: si tratta di un algoritmo capace di prevedere dove e quando è più elevata la possibilità che si verifichino attività di spaccio di sostanze stupefacenti, sulla base dell’esame di una serie di fattori spaziali e ambientali tradizionalmente in grado di favorirne la commissione, quali ad esempio la vicinanza a locali notturni, a stazioni ferroviarie, a fermate di mezzi pubblici, scarsa presenza di luminarie stradali, orari notturni etc. Tale modus operandi, dunque, si è rivelato capace di consentire di disegnare una vera e propria mappatura delle aree urbane dove è maggiore il rischio di spaccio di sostanze stupefacenti, e di programmare e porre in essere con adeguato anticipo interventi volti ad impedire la commissione del fatto criminoso. Un altro esempio di software di polizia predittiva che si ispira alle acquisizioni della criminologia ambientale potrebbe corrispondere anche a un dispositivo attualmente utilizzato dalla Polizia di Stato italiana: il cosiddetto “X-LAW”, messo a punto in origine dalla questura di Napoli. Anche il funzionamento di tale algoritmo si basa sostanzialmente sulla rielaborazione di un’ingente mole di dati e informazioni derivanti dalle denunce inoltrate nel tempo alla Polizia di Stato, volta a far emergere coincidenze, correlazioni e fattori ricorrenti che l’agente umano potrebbe non essere altrettanto capace di osservare. In questo modo diviene possibile disegnare una mappa del territorio che evidenzi le zone e gli orari “caldi”, dove è più elevato il rischio di commissione di attività criminali, e sulla base delle indicazioni che risultano, predisporre le forze dell’ordine al fine di impedire che tali attività vengano portate a compimento, cogliendo in flagranza i relativi autori.

È riconducibile, invece, alla categoria dei software ispirati al concetto di crime linking il cosiddetto “Software Keycrime”, elaborato presso la Questura di Milano, non dissimile da altri software della medesima tipologia elaborati e attualmente in uso in Germania, Inghilterra e Stati Uniti. Mentre i sistemi di individuazione degli hotspots si basano sulla individuazione delle “zone calde”, al contrario i sistemi di crime linking si concentrano sul profilo della persona potenziale autore di un reato. Il funzionamento di tali algoritmi, dunque, individua il proprio nucleo nella capacità di disegnare il profilo criminale del possibile autore di una serie di fatti criminosi, al fine di ipotizzarne la mossa successiva.

Eppure di tali sistemi di polizia predittiva, di cui si è tentato di offrire un sintetico quadro descrittivo, se da un lato siano considerevoli ed innegabili i vantaggi sul piano della prevenzione del crimine – vantaggi che col tempo sono certamente destinati a crescere e ad intensificarsi – dall’altro lato non se ne possono trascurare le perplessità che scaturiscono dal loro utilizzo, al fine di una loro efficiente ed allo stesso tempo sicura attuazione.

In primo luogo bisogna osservare che tale primo impiego dei sistemi di Intelligenza Artificiale, ad oggi, non è stato affatto disciplinato in nessuno Stato, ragione per la quale il loro uso rimane affidato esclusivamente al buon senso, alla sensibilità e all’esperienza degli operatori di polizia, e del tutto svincolato da ferree disposizioni scritte.

In secondo luogo vanno considerate anche le implicazioni circa la tutela della privacy, in considerazione del fatto che il funzionamento della maggior parte di tali sistemi di polizia predittiva si basa principalmente su algoritmi in grado di acquisire e rielaborare notevoli quantità di informazioni e dati personali, nel necessario rispetto del divieto di discriminazione nell’analisi delle caratteristiche etniche, somatiche, religiose e sociali.

Ancora, per quanto attiene al loro meccanismo di funzionamento, tali software si autoalimentano con i dati che essi stessi rilevano, emergendo così il rischio che il loro utilizzo divenga controproducente rispetto agli obiettivi da raggiungere e che si inneschino irrimediabili circoli viziosi. In altre parole, se il software individua una determinata area come “zona calda”, con conseguente intensificazione dello spiegamento delle forze dell’ordine, sorge il rischio che da quel momento in poi i sistemi di polizia predittiva continuino a qualificare quella regione come zona calda non già in ragione dell’effettiva verificazione di fatti criminosi, bensì in ragione dell’intensificazione dei controlli e dei pattugliamenti e del fatto che quell’area sia stata qualificata in origine come zona a rischio dal software stesso.

Inoltre va rilevato come tali sistemi di polizia predittiva conducano alla prevenzione dei reati attraverso la militarizzazione delle aree emerse come “zone calde” — tramite cioè il pattugliamento e l’intervento attivo delle forze di polizia – lì dove sarebbe invece più auspicabile condurre la battaglia al crimine sul piano sociale, attraverso cioè un’azione che si rivolga ai fattori criminogeni, quali i fattori economici, ambientali, individuali etc.

Da ultimo, bisogna tener conto che la maggior parte di questi software, in quanto messi a punto da aziende private, sono coperti dal segreto industriale e commerciale, ragione per la quale non sarebbe possibile comprendere a fondo il meccanismo del relativo funzionamento, con conseguente pregiudizio delle esigenze di trasparenza e pubblicità.

Naturalmente, tali perplessità legate all’utilizzo dei sistemi di polizia predittiva non dovrebbero rappresentare una ragione per arrestarsi dinanzi al progresso, quanto invece un incentivo regolamentarne l’impiego ed a predisporre una specifica risposta giuridica alle questioni legali che l’uso di tali algoritmi può presentare.

**Gli “Automated decision system” e la “Carta etica europea per l’uso dell’Intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti”.**

Il secondo ambito coinvolto dalla diffusione delle Intelligenze artificiali concerne l’impiego di algoritmi decisionali volti a consentire l’affiancamento, o addirittura la sostituzione, del giudice-macchina rispetto al giudice-uomo.

Tali sistemi di Intelligenza Artificiale, definiti “Automated decision system”, sono in fase di forte crescita e diffusione sia nel pubblico che nel privato e, in ragione della loro capacità di acquisire in tempi rapidi una molteplicità di dati da innumerevoli fonti quali Codici, banche-dati giurisprudenziali, legislative e raccolte di precedenti, si sono rivelati in grado di assumere decisioni volte a prevenire e comporre le controversie tra i privati.

Si tratta dunque, anche da questo diverso angolo visuale, di una potenziale rivoluzione avente ad oggetto l’introduzione di veri e propri meccanismi alternativi di risoluzione delle controversie che rispetto al sistema giuridico tradizionale comportano vantaggi di non poco momento nell’economia globale di un procedimento giudiziario, quali la riduzione dei tempi e rilevanti risparmi di spesa per le parti in causa. Ciononostante, gli “Automated decision system” sono oggi adoperati quasi esclusivamente nell’ambito di alcune questioni civili, concernenti ad esempio il risarcimento dei danni, danni da prodotto, pratiche assicurative, mancando ancora un loro utilizzo in ambito penale, in ragione, probabilmente, della maggiore delicatezza e rilevanza di tale settore. Naturalmente l’attuale realtà non esclude affatto che l’incalzante progresso tecnologico possa condurre in tempi brevi ad un ruolo decisionale attivo di tali software anche nelle cause penali. Ed infatti, proprio la prospettiva di una diffusione di algoritmi decisionali anche in materia penale ha destato l’attenzione e suscitato le preoccupazioni del Consiglio d’Europa, la cui commissione per l’efficacia della giustizia (CEPEJ), lo scorso 3 dicembre 2018, ha redatto la cosiddetta “Carta etica europea per l’uso dell’Intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti”. Si è trattato di un epocale momento di svolta nel mondo dell’interazione tra Intelligenza Artificiale e Diritto: per la prima volta, sebbene solo limitatamente al contesto di tale impiego dei sistemi di Intelligenza artificiale, si è preso atto del loro smisurato potenziale e del ruolo attivo che sono inevitabilmente destinati a svolgere nella società moderna, e sulla base di questo, si è tentato di cristallizzare alcune fondamentali linee guida a cui dovranno necessariamente attenersi i responsabili dei progetti aventi ad oggetto lo sviluppo e l’uso delle IA.

Tale Carta etica europea ha avuto il merito di affermare, dunque, anche nell’ambito dell’utilizzo dei sistemi di Intelligenza artificiale, principi basilari quali il rispetto dei assicurative, uffici e studi legali, avvocati e privati” di tali sistemi di Intelligenza Artificiale in questo campo necessita la predeterminazione di una serie innumerevole di criteri che il software possa prendere in considerazione, quali l’età, il sesso, il livello di scolarizzazione, la situazione sociale, familiare e lavorativa, l’origine etnica, i precedenti penali, luoghi e persone frequentate, vicinanza ad ambienti criminali, dipendenza da alcol o da sostanze stupefacenti, infermità mentale etc.

Pertanto, la rilevazione di uno o più di tali fattori da parte dell’algoritmo predittivo, in capo al medesimo individuo, determinerà un grado più o meno elevato di pericolosità criminale ovvero del rischio di commissione di un nuovo reato, sulla base di un approccio di tipo attuariale o statistico.

Inoltre, gli indicatori evidenziati possono essere suddivisi in tre micro-categorie: vi sono in primo luogo i fattori statici, ovvero non modificabili, come il sesso e l’origine etnica; in secondo luogo i fattori dinamici stabili, modificabili attraverso un trattamento terapeutico, come l’infermità mentale; in terzo luogo i fattori suscettibili di modificarsi con maggiore rapidità, come l’assunzione di alcol o sostanze stupefacenti.

Senza dubbio gli algoritmi predittivi nell’ambito di tale funzione offrono l’innegabile vantaggio di compiere valutazioni accurate e in maniera piuttosto rapida circa la pericolosità criminale di singoli soggetti. Ciononostante, il loro raggio d’impiego si limita oggi esclusivamente agli Stati Uniti, mentre l’Europa si dimostra restia ad aprirsi ad un loro utilizzo nei tribunali in ragione di diverse difficoltà non trascurabili che inevitabilmente fanno da contraltare agli indubbi meriti.

**Le criticità dei sistemi predittivi.**

Una prima motivazione la si può rinvenire nell’art. 22 del nuovo Regolamento europeo in materia di protezione dei dati personali, introdotto nel maggio del 2018 in attuazione della direttiva comunitaria 95/46/CE, in cui si afferma il principio in virtù del quale nessuna persona può venire sottoposta ad una decisione giurisdizionale che sia fondata esclusivamente su un trattamento automatizzato di dati. Ma lasciando per un momento da parte tale norma di carattere generale, sono altre e numerose le difficoltà pratiche derivanti dall’impiego degli algoritmi predittivi: innanzitutto, essendo il funzionamento di tali software caratterizzato da un approccio di tipo attuariale e statistico – nella misura in cui essi sono predisposti in maniera tale da leggere e rielaborare una serie di innumerevoli fattori di rischio predeterminati – risulta fortemente compromessa l’esigenza di una valutazione individualizzata della dunque, sia possibile verificare il funzionamento interno dei software impiegati nell’ambito della valutazione della pericolosità criminale e l’attendibilità scientifica dei risultati da essi forniti, parrebbe opportuno intervenire a livello legislativo su due fronti: in primo luogo rendere accessibili al pubblico – giudice, parti processuali, e la generalità degli utenti – le informazioni relative alla struttura del software e che nessun segreto industriale possa essere opposto alla richiesta di analizzarne l’architettura; in secondo luogo garantire l’indipendenza dell’algoritmo dal suo stesso programmatore al fine di scongiurare il rischio che ne venga compromessa l’imparzialità, nella medesima ottica dell’esigenza dell’imparzialità del giudice.

Inoltre, non bisogna nemmeno trascurare la necessità di formare l’intero personale giudiziario – dai giudici alle cancellerie – affinché sia in grado di servirsi dei sistemi di Intelligenza artificiale, allestendo corsi di aggiornamento che inevitabilmente richiederebbero ingenti investimenti da parte del Governo.

Ulteriori criticità possono poi derivare dalla circostanza che gli algoritmi predittivi, se da un lato sono strutturati in maniera tale da essere in grado di attingere a precedenti giurisprudenziali per elaborare la risoluzione del caso loro sottoposto, dall’altro lato non sempre si rivelano capaci di comprendere che la controversia in esame non necessariamente rappresenta un “caso simile” ai precedenti giurisdizionali analizzati. Tale andamento, pericolosamente insensibile alle singolarità di ogni caso che arricchiscono la giurisprudenza, rischia di favorire quindi una cristallizzazione della giurisprudenza stessa.

L’ultimo effetto che proprio tali problematicità minacciano di produrre sul piano etico lo si potrebbe riconoscere nella potenziale deresponsabilizzazione dei giudici: è concreto il rischio, tanto nell’ambito della valutazione della pericolosità criminale da parte degli algoritmi predittivi, quanto nel campo d’impiego dei sistemi di Intelligenza artificiale a fini decisionali – specie quando la macchina robotica si limita ad affiancare l’operatore umano anziché sostituirlo – che i giudicanti si appiattiscano sull’ipotesi di soluzione offerta dal software, delegando ad esso le proprie responsabilità e di fatto abdicando alla propria funzione.

Ad ogni modo, specialmente negli Stati Uniti, tali perplessità non hanno affatto arrestato l’avanzata dei software per la valutazione della pericolosità criminale – di cui infatti viene fatto un largo uso nel sistema americano di giustizia penale – ed anzi si sono in parte manifestate nell’impiego dell’algoritmo predittivo di gran lunga più famoso, il “Compas” (“Correctional Offender Management Profiling for Alternative Sanctions”). Tale software, elaborato per la prima volta nel 1998, viene oggi utilizzato in quasi tutti i Paesi degli Stati Uniti non solamente per la valutazione del pericolo di recidiva, ma anche per individuare le esigenze trattamentali del singolo individuo volte a ridurre il rischio di commissione di nuovi reati. Tra i principali fattori che vengono esaminati dal Compas figurano precedenti penali, frequentazioni criminali, abuso di alcol e sostanze stupefacenti, difficoltà economiche, contesto sociale, e sebbene tra essi non compaia alcun riferimento all’origine etnica o alle caratteristiche somatiche, alcuni studi(J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, Machine Bias, in www.propublica.org, 23 maggio 2016) avrebbero rilevato come le previsioni formulate dal software risultassero inaffidabili in quanto discriminatorie, nella misura in cui favorissero gli imputati bianchi rispetto agli imputati di colore calcolando, su base razziale, un minor rischio di recidiva per i primi ed un maggior rischio per i secondi.

Oltre al difetto della discriminazione razziale, ulteriori ricerche ( J. DRESSEL, H. FARID, The accuracy, fairness, and limits of predicting recidivism, 30 marzo 2018.) hanno dimostrato che le previsioni offerte dal Compas sulla base dell’analisi di una molteplicità di indicatori risulterebbero equivalenti sia a quelle che potrebbero essere fornite anche da un soggetto completamente privo di conoscenze e competenze specifiche in materia, sia a quelle che lo stesso Compas potrebbe fornire sulla base dell’analisi di soli due fattori, mettendo definitivamente in discussione l’effettiva utilità di tale algoritmo predittivo. A molte di queste critiche è stata data voce in particolare modo nel cosiddetto caso Loomis ( Wisconsin S.C. State V. Loomis, 2016) , sebbene in relazione all’impiego del Compas a fini decisionali nell’ambito della commisurazione della pena piuttosto che a fini predittivi della valutazione del rischio di recidiva: Loomis, condannato dalla Corte locale, aveva impugnato la sentenza dinanzi alla Corte Suprema del Wisconsin per ricorrere contro il contributo offerto in primo grado dal Compas, di cui veniva contestato il pregiudizio razziale e la violazione del principio di trasparenza, in ragione del segreto industriale che ne copre il meccanismo di funzionamento. La Corte Suprema del Wisconsin, dunque, aveva evidenziato in termini di critica l’impossibilità di una conoscenza piena a approfondita del funzionamento interno del software, la violazione dell’esigenza di una valutazione individualizzata della pericolosità sociale, ed infine il rischio di discriminazione razziale, nella misura in cui fosse stato sovrastimato il rischio di recidiva in relazione a talune minoranze etniche. Nel caso di specie però la Corte Suprema respinse il ricorso dell’imputato Loomis in considerazione del fatto che le valutazioni formulate dal Compas non rimasero autonome e determinanti, ma furono supportate e avvalorate dal giudice-uomo, ma ciononostante rimangono decisive, per i futuri sviluppi di tali sistemi di Intelligenza artificiale e per il relativo impiego, le critiche rivolte dalla Corte Suprema nei confronti degli algoritmi predittivi.

Maggiore successo ha riscontrato invece il “Public Safety Assessment”, un algoritmo predittivo della valutazione della pericolosità criminale elaborato nel New Jersey e adoperato anche in altri Paesi degli Stati Uniti. Si tratta di un software in grado di confrontare gli indicatori del rischio di recidiva riscontrati nell’imputato sottoposto a valutazione con una molteplicità di altri casi provenienti da altre giurisdizioni del Paese, al fine di attribuire all’imputato in esame un punteggio relativo al rischio di commissione di nuovi reati, in scala da uno a sei, il più oggettivo possibile, e che tenga conto degli altri casi simili negli Stati Uniti. Il Public Safety Assessment ha consentito di conseguire notevoli vantaggi permettendo un’impennata delle scarcerazioni durante la celebrazione del processo sulla base di un’accurata e neutrale valutazione della pericolosità criminale ed in particolar modo del rischio di recidiva, senza neppure il pagamento della cauzione per gli imputati non abbienti.

L’impetuosa diffusione di tali sistemi algoritmici, nonostante le perplessità legate al loro utilizzo, potrebbe rendere lecito domandarsi se il progresso tecnologico non sia destinato a conquistare anche il mondo del Diritto, a rimediare alla complessità, ai costi e ai tempi connessi ai processi condotti dall’operatore umano, e se il prezzo di tutto questo non debba essere proprio una parziale rinuncia ad alcuni dei principi fondamentali. Ma gli evidenti meriti, in questo come negli altri ambiti d’impiego dei sistemi di Intelligenza artificiale, dovrebbero sollecitare il legislatore a regolamentarne la sfera operativa, al fine di impedirne un uso sconsiderato aperto ad ogni compromissione delle garanzie fondamentali, e senza ostacolare il progresso e l’innovazione.

**Robot e reati.**

Un quarto ed ultimo ambito di cui potrebbe essere protagonista nella società moderna un sistema di Intelligenza artificiale concerne le possibili ipotesi di coinvolgimento di una macchina robotica nella commissione di un reato, in qualità di strumento, vittima, o infine autore dell’illecito.

Per quanto riguarda l’ipotesi che vede un software, un algoritmo, o un dispositivo robotico quale strumento di un reato – si tratta dell’eventualità più frequente, o per lo meno più plausibile –, vanno indagate a fondo le enormi potenzialità che un sistema di Intelligenza artificiale potrebbe offrire nella commissione di un illecito penale. Sono in effetti molteplici, e si avviano a divenire sempre maggiori, le condotte criminali che possono essere agevolate dallo sfruttamento di sistemi algoritmici: frode informatica, cyber-terrorismo, cyber-laundering, traffico di sostanze stupefacenti e di altri prodotti illeciti, traffico di esseri umani, violazione della privacy, violazione della proprietà intellettuale e industriale, trattamento illecito di dati personali, diffamazione online, divulgazione di fake news, per concludere, infine, con l’ipotesi che più si avvicina alla possibilità che vede un sistema d’Intelligenza artificiale nelle vesti dell’autore del reato, consistente nella progettazione, da parte del programmatore umano, di una macchina robotica volta alla commissione di un reato nell’interesse e per conto dell’uomo. Da segnalare, ancora, le condotte di manipolazione abusiva del mercato, che possono essere commesse dall’uomo avvalendosi di accurati sistemi informatici, in grado non soltanto di eseguire concretamente le transazioni finanziarie, ma anche di adottare autonomamente la decisione di porle in essere grazie ad un algoritmo capace di comparare rapidamente numerosissime variabili: si tratta dei cosiddetti “High Frequency Traders” (HFT), in grado di eseguire migliaia di operazioni al secondo. Dunque, l’impiego di questi sofisticati programmi informatici per scopi criminali può provocare profonde oscillazioni dei prezzi di mercato, e senza che a tali variazioni corrisponda un effettivo mutamento del valore sostanziale del titolo in questione, integrandosi così la condotta illecita di manipolazione abusiva del mercato. Come anticipato, l’uomo si trova in balia della tecnologia, e agli incommensurabili benefici e vantaggi che questa offre in ogni ambito della vita quotidiana, non potrebbero non corrispondere altrettanti rischi. In questo scenario, pertanto, è opportuno che il Diritto stia al passo con il progresso e l’innovazione tecnologica, al fine di non lasciare l’uomo sguarnito di ogni protezione, essendo i presidi tradizionali senz’altro idonei a offrire una tutela giuridica dinanzi alle illecite condotte umane, come il sistema penale; ma quando un sistema di Intelligenza artificiale raggiunge un livello di autonomia e di capacità di apprendimento tale da essere in grado di adottare decisioni e comportamenti in maniera del tutto indipendente dal proprio programmatore? In capo a chi potrà riconoscersi la responsabilità delle condotte criminose commesse da un dispositivo robotico privo di comandi, quando il suo utilizzo non sia quindi riconducibile ad un utilizzatore umano?”.

*Il testo integrale della tesi di Antonio Carlo Oliveri del Castillo è in via di pubblicazione*