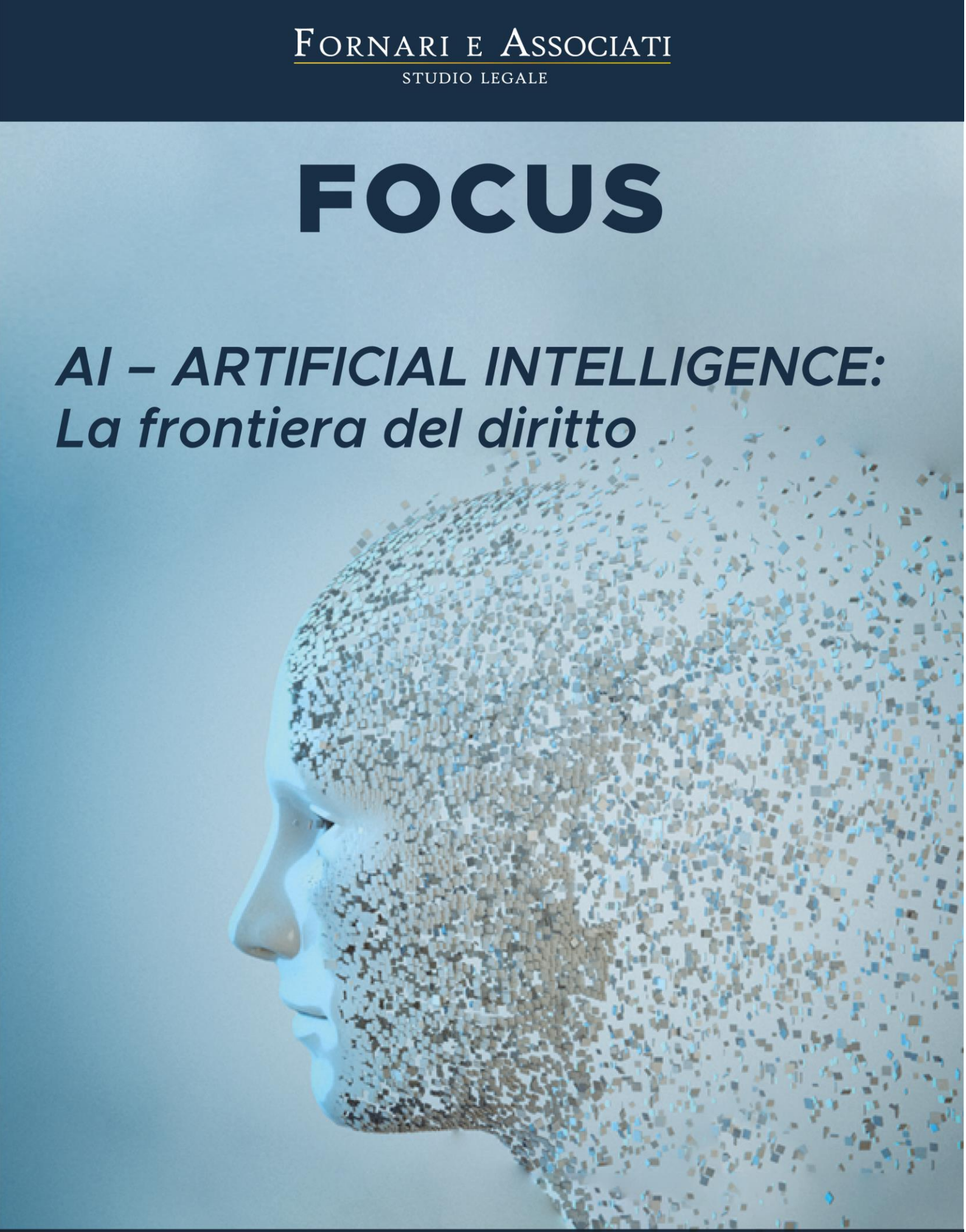
FORNARI E ASSOCIATI
STUDIO LEGALE

FOCUS

AI – ARTIFICIAL INTELLIGENCE: La frontiera del diritto



fornarieassociati.com
Milano • Roma
Ottobre 2020

AI – ARTIFICIAL INTELLIGENCE: LA FRONTIERA DEL DIRITTO

Strutturare un *corpus* normativo in grado di contenere al suo interno, e così governare, il complesso e variegato fenomeno dell'intelligenza artificiale è una delle sfide – se non la sfida – del prossimo futuro.

Ad oggi, il percorso che conduce alla futura regolamentazione giuridica dell'intelligenza artificiale sembra essere costellato da molteplici dubbi ed una sola certezza.

La certezza, ormai consolidata, è quella che vede l'intelligenza artificiale diffondersi a macchia d'olio fino a permeare, nel prossimo futuro, le dinamiche socio-economiche della civiltà post-moderna. È altrettanto certo che non si tratterà di un processo lento – bensì rapido, pure a fronte della sua profondità – ed irreversibile: la diffusione lenta e progressiva è appannaggio di tecnologie ormai datate, così come la TV, la radio, l'elettricità, l'automobile, le quali hanno impiegato più 50 anni per raggiungere i 50 milioni di utenti, lasciando così al giurista il tempo necessario per assorbire le innovazioni da queste introdotte ed improntare regole volte ad organizzare la società intorno ad esse. Al contrario, *Twitter* ha impiegato meno di 3 anni per raggiungere i 50 milioni di utenti e *Facebook* meno di 2, rischiando così di lasciare il giurista – a tecnologia già inserita nel circuito economico-sociale – paralizzato da uno '*shock di modernità*'.

C'è da scommettere che il *trend* di diffusione seguito dall'intelligenza artificiale si porrà nella scia di *Twitter* o *Facebook*, piuttosto che in quella della TV o dell'automobile, con la differenza che – dinnanzi ai nodi antropologici, filosofici, sociali, economici, giuridici e psicologici posti dall'AI – l'eventuale inerzia o scompostezza del giurista potrebbe condurre a conseguenze drammatiche. Drammatiche al punto da porre in dubbio l'idoneità dell'armamentario giuridico a disciplinare il progresso tecnologico in modo economicamente efficiente ed ordinato rispetto ad una scala assiologica di valori, introducendo così una riflessione su quella che è stata efficacemente definita la *'fine del diritto'*.

Il presente contributo illustra i percorsi di indagine attualmente tracciabili in materia di interferenze reciproche tra intelligenza artificiale e diritto penale: dai nodi irrisolti circa l'allocazione della responsabilità penale derivante da fatti di reato commessi dall'intelligenza artificiale (*machina delinquere potest?*) all'implementazione di applicativi AI all'interno del processo penale (*A.I., Procedimento Penale e Law-Enforcement).*

Nel regime di incertezza che governa la materia, la conclusione non può che essere unica: nell'anticamera del futuro, un serrato confronto tra scienza giuridica e scienza cibernetica non è postergabile.

I. *MACHINA DELINQUERE POTEST*?

Come in una scatola cinese, la risposta all'interrogativo contiene in sé un'infinità di ulteriori punti di domanda.

Difetta, innanzitutto, una definizione di *'intelligenza artificiale'*; difetta, quantomeno, una definizione sintetica, esaustiva ed universalmente condivisa da poter elevare a descrizione normativa del fenomeno.

Il termine '*artificial intelligence*' – coniato nel 1955 da John McCarthy, uno dei padri fondatori dell'AI – ha incamerato nel tempo il vorticoso processo di evoluzione tecnologica snodatosi in via esponenziale a partire dagli anni '60 del secolo scorso, finendo oggi con il racchiudere al suo interno una serie di dispositivi *hardware* con capacità computazionali in allora soltanto immaginabili con fervida fantasia.

Un tentativo definitorio della nozione di '*intelligenza artificiale*' indicizzato allo sviluppo tecnologico contemporaneo è stato esperito da un gruppo di 52 esperti appositamente nominati dalla Commissione Europea, i quali – nell'ambito della Comunicazione del 2018 denominata *"Una definizione di AI: principali capacità e discipline scientifiche"* – si sono espressi nei seguenti termini: *"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****"*.

È opportuno calare la definizione teorica nell'alveo della realtà concreta per materializzare dinnanzi ai nostri occhi l'impatto d'urto che l'AI produrrà in seno nostra società nel corso del prossimo futuro: *robot* – non necessariamente umanoidi – autonomi nell'apprendimento della realtà circostante e nell'esecuzione dei compiti loro affidati (potenzialmente, dalla pulizia delle strade urbane ad operazioni chirurgiche a cuore aperto); automobili a guida autonoma; droni da guerra privi di conduzione o controllo umano; *high-frequencies trader* operanti in autonomia sui mercati finanziari; applicazioni IOT (*Internet of Things*).

La linfa vitale che alimenta questi dispositivi – e gli algoritmi che li governano – sono i nostri dati. Ogni persona sul pianeta genera, ogni secondo, 1,7 megabyte di dati. Ogni giorno l'umanità genera 2,5 quintilioni di gigabyte di dati, dove un quintilione è pari ad un miliardo di triliardi, la base dieci elevata alla trentesima. Il 90% dei dati conservati nella storia dell'umanità è stato generato nel corso degli ultimi due anni. In assenza di tale vorticosa produzione di dati, l'intelligenza artificiale non avrebbe modo di esistere, poiché verrebbe a mancare la base cognitiva che la macchina sottopone a computazione ai fini dello sviluppo e dell'evoluzione della propria *intelligenza*. E la AI – come ogni essere autonomo – sarà in grado di produrre a sua volta la linfa di cui si nutre: i dati. Non solo, infatti, l'interazione tra umani e macchine genererà nuovi *cluster* di dati (*machine to human* e *human to machine*), bensì la stessa interazione tra macchine sarà fattore da cui prolifererà la produzione di nuovi dati (*machine to machine*).

Secondo fattore determinate per la sussistenza dell'intelligenza artificiale è l'evoluzione tecnologica lato *hardware* e lato *software*: il progresso ha finalmente indotto sul mercato la presenza – a prezzi economicamente utili e a dimensioni fisiche ridotte – di dispositivi *hardware* in grado di supportare il processo di

assimilazione dati (*e.g.* sensori audiovisivi, spaziali, biometrici) ed il processo di computazione (*e.g.*algoritmi) tipico dell'AI. Il terreno è dunque pronto – non solo per l'esistenza di *software* in grado di supportare algoritmi contemplanti processi di computazione AI – ma altresì per introdurre tali *software* in strutture fisiche *hardware* dalle più disparate morfologie e funzionalità.

E il contributo umano si arresta a questa soglia. Fornita la base dati e dotata la macchina di un algoritmo contemplante funzionalità AI, la stessa macchina sarà in grado di analizzare in autonomia le informazioni presenti nella realtà circostante e di computare quest'ultime secondo processi non rigidamente tracciati o alternativamente previsti dal programmatore umano, bensì creativamente strutturati dalla stessa macchina poiché ritenuti efficienti ai fini del perseguimento dell'obiettivo preposto. Il comportamento della macchina non è dunque assistito da pre-determinismo. Al contrario, il sistema AI è in grado: *(i)* in un primo momento, di compilare la propria base cognitiva attingendo dalla sconfinata mole di dati presenti in *cloud*; di apprendere ulteriori informazioni dalla realtà circostante tramite sensori ambientali; di comunicare con altre macchine operanti nello stesso o in un diverso ambiente; *(ii)* in un secondo momento: di computare la piattaforma cognitiva attraverso processi autonomamente strutturati; di agire sulla scorta di tali processi; di imparare dall'esperienza e modificare la propria condotta futura sulla scorta dei propri errori; di accrescere in via progressiva ed esponenziale il proprio processo di apprendimento e l'efficacia-efficienza della propria azione. Tutto ciò in assenza di qualsiasi intervento umano. Non solo: il processo che conduce dall'*input* all'*output* – proprio perché strutturato in autonomia dalla macchina – resta per vasti tratti non intellegibile dall'intelligenza umana, con la conseguenza che il percorso computazionale che conduce la macchina all'azione non risulta, oltre che prevedibile *ex ante*, nemmeno ricostruibile *ex post* dall'uomo.

I *software* che conferiscono autonomia computazionale (e quindi decisionale) alle macchine sono denominati *black box algorithms* e traggono il loro nome dalla *scatola nera* in cui l'AI si trova ad agire nell'esercizio della sua autonomia; i processi computazionali operati in autonomia dalle macchine vengono invece denominati processi di *machine learning* e *deep learning*. Trattasi, senza dubbio, del tratto singolare e maggiormente qualificante il fenomeno AI.

È nella scatola nera – non intellegibile alla mente umana – dell'AI che si sviluppano le potenzialità del nostro futuro: dalla possibilità di avere automobili a guida autonoma sulle strade alla possibilità di avere *trader* informatici strutturanti ottimali strategie di investimento finanziario, e via dicendo.

È in questa scatola nera, tuttavia, che può trovare formazione un processo rappresentativo e deliberativo – di matrice non umana – sfociante nella consumazione di un fatto di reato.

Nel maggio 2010, a Wall Street, *high frequencies traders* operanti in Borsa hanno causato un *flash crash* nel quale si è improvvisamente vanificata capitalizzazione per 1 trilione di dollari. Nel 2013, sempre a Wall Street, altri *high frequencies traders* hanno processato la *fake news* di un falso attentato a Barack Obama; prevedendo un correlata flessione a ribasso dei titoli quotati, hanno innescato una massiccia liquidazione di investimenti, causando così il crollo dell'indice di borsa. Nel luglio 2016, a Dallas, un cecchino è stato annientato, tramite uccisione, da un drone agente in logica AI. Nel marzo 2018, in Arizona,

una automobile a guida autonoma ha travolto, uccidendolo, un pedone. Sono solo sporadici esempi. Eppure, una misura di ciò che potrà essere la nostra futura convivenza con l'AI, dal momento che i più recenti studi di settore prevedono entro il 2030 una diffusione a macchia d'olio dell'AI almeno nei seguenti settori: trasporti, finanza, domotica, sanità, istruzione e sicurezza pubblica.

A fronte di ciò, il nodo giuridico da sciogliere è quello della corretta allocazione della responsabilità penale per i fatti commessi dalle macchine. La sfida è complessa e la posta in gioco elevata: il vuoto di tutela penale per fatti di reato commessi dall'AI.

I.1. *MACHINA DELINQUERE POTEST*.

Si ponga il caso di un applicativo AI operante in via legittima nella società contemporanea poiché selezionato dal legislatore nell'ambito del perimetro di *rischio consentito* tollerato dall'ordinamento giuridico e, come tale, autorizzato. Si ponga altresì il caso che siffatto applicativo AI non venga in alcun modo abusato o alterato nel suo funzionamento dall'essere umano, ma che sviluppi un genuino processo di *decision making* sulla scorta dei meccanismi di *deep* e *machine learning* tarati su *black box algorithms* e, come tale, un processo di *decision making* non prevedibile *ex ante* né controllabile *ex post* dall'essere umano. Si ponga il caso, infine, che tale processo di *decision making* sfoci accidentalmente nella commissione di un fatto di reato ad opera della macchina: l'errore mortale nel corso di un'operazione condotta da un chirurgo *robot*, l'investimento mortale di un pedone da parte di una automobile senza conducente, il *market abuse* commesso da un *high frequencies trader*.

Chi potrà essere ritenuto penalmente responsabile di un tale fatto di reato e, quindi, ad esempio, dell'evento morte del paziente o del pedone? Secondo alcuni, potrà esserlo direttamente la macchina alimentata da AI.

La risposta affermativa all'interrogativo *'machina delinquere potest?'* proviene dalla dottrina penalistica israeliana e, in particolar modo, porta la firma di Gabriel Hallevy. *"What else is needed?"* si chiede Gabriel Hallevy alla fine del suo *paper* "*Dangerous Robots*", ossia, che cos'altro manca affinché si possa addivenire al riconoscimento di responsabilità penale in capo agli applicativi AI? A detta sua, nulla. L'Autore sostiene infatti che il brocardo *machina delinquere non potest* sia dotato di una solidità soltanto apparente, destinata a vacillare una volta che la materiale iniezione di AI nel circuito economico-sociale avrà evidenziato l'esigenza pragmatica di addivenire all'accertamento di responsabilità penale in capo alle macchine. Ciò è già avvenuto – sostiene sempre l'Autore con un suggestivo parallelismo – quando il brocardo *societas delinquere non potest* coniato dalla dottrina penalista tedesca sul finire dell'800 si è sfaldato per il tramite dell'introduzione, in pressoché tutti gli ordinamenti giuridici occidentali, di una responsabilità penale o para-penale in capo ad enti e persone giuridiche. Una responsabilità che è stata introdotta nell'ordinamento giuridico italiano per mano del celebre d.lgs. n. 231/01 e che negli anni – divenuta parte integrante ed irrinunciabile dell'ordinamento giuridico – è stata oggetto di continue espansioni.

Il parallelismo formulato da Hallevy è suggestivo, ma ciò non può eclissare i molteplici ostacoli concettuali al riconoscimento di una responsabilità penale in capo alle macchine. L'attribuzione di una responsabilità (para)-penale in capo alle persone giuridiche (società, *in*

primis) è stata infatti osteggiata per decenni sulla scorta del fatto che la sanzione penale sarebbe stata del tutto inutile nei loro confronti, poiché in capo ad una società l'irrogazione di una pena non avrebbe potuto produrre alcuna finalità rieducativa o preventiva. L'utilità della sanzione penale si sarebbe quindi inevitabilmente infranta sull'assenza di un corpo fisico da privare di libertà e di una coscienza da ricondurre a rettitudine (*"no body to kick, no soul to damn"*). L'ostacolo concettuale è stato dunque aggirato decenni più tardi prevedendo sanzioni penali – sì calate in via formale in capo alla società – ma affliggenti in via sostanziale gli organi di amministrazione di questa dinnanzi agli *stakeholders* che sono chiamati a rappresentare, tramite *targettizzazione* del fulcro intorno a cui si condensano gli interessi economici di tali persone fisiche: il bilancio e le casse della società o dell'ente.

Tale punizione dell'umano attraverso l'inumano, tuttavia, non pare potersi agevolmente replicare per quanto riguarda la criminalizzazione degli applicativi AI.

Al contrario di enti e società, che ne sono state storicamente dotate, difetterebbe in primo luogo in capo alle macchine una soggettività giuridica. La sanzione penale in capo alle macchine postulerebbe dunque la teorizzazione di un nuovo *'diritto penale delle cose'*, all'interno del quale macchine AI dovrebbero acquisire lo *status* di *"soggetti del diritto"*: se non dei veri e propri agenti sul piano penale al pari degli esseri umani, quantomeno degli *'attanti'* la cui condotta sia in grado di elevarsi a presupposto per l'applicazione di sanzioni penali.

Ciò nonostante, il fatto di reato risulterebbe comunque composto da un pregnante elemento soggettivo governato dal principio di colpevolezza. Si rivelerebbe dunque imprescindibile comprendere la possibile attribuzione alle macchine della *'capacità di intendere e di volere'*, criterio fondante l'imputabilità soggettiva di responsabilità penale. Inoltre, imprescindibile sarebbe la comprensione circa la possibilità della macchina di maturare un istante rappresentativo ed un istante volitivo da porre a sostegno delle proprie decisioni e conseguenti azioni, così da integrare i presupposti che sorreggono l'attribuzione di responsabilità penale a titolo doloso. In questo senso, soltanto un confronto tra scienza giuridica, scienza cibernetica e neuroscienze potrebbe restituire una fotografia di insieme delle analogie e differenze che sorreggono il percorso deliberativo degli esseri umani e delle macchine, così da fornire una indicazione circa la possibilità di traslare *in toto* alle macchine i principi di attribuzione della responsabilità soggettiva umana, oppure circa la necessità di strutturare un nuovo *'diritto penale delle cose'* in cui ri-disegnare ad immagine e somiglianza dell'AI i principi di rappresentazione e volizione della condotta illecita.

E ancora, l'interrogativo che rimarrebbe ad ogni modo irrisolto sarebbe il seguente: perché infliggere una pena alla macchina? In seno al diritto penale moderno, la comminazione ed applicazione di sanzioni penali si giustifica soltanto alla luce dell'effetto dissuasivo che la comminazione della sanzione può produrre sulla generalità dei consociati (finalità general-preventiva), dell'effetto dissuasivo che l'applicazione della sanzione può avere sul singolo condannato (finalità special-preventiva) e, soprattutto, dell'effetto ri-educativo e di ri-socializzazione che l'esecuzione della pena può indurre in capo al reo (finalità rieducativa).

Nulla di tutto ciò sembra replicabile nel mondo delle macchine. Poiché essa è priva di una coscienza etica generalizzata e condivisa tra

unità robotiche, l'applicazione della pena in capo ad una singola macchina non pare in grado di suscitare un monito di deterrenza nei confronti dell'IA complessivamente considerata. Poiché essa è priva di emozioni ed empatia, l'applicazione della pena in capo ad una singola macchina non pare in grado di orientare la sua condotta futura, né, tantomeno, di svolgere alcuna funzione pedagogica.

Di fatto, dunque, l'applicazione di sanzione penale ad una macchina – nella forma della sospensione dalla circolazione o della distruzione, piuttosto che del lavoro di pubblica utilità – finirebbe soltanto con il soddisfare l'umano sentimento di giustizia delle persone offese o danneggiate dal reato; una finalità, quest'ultima, assai labile e tra l'altro suscettibile di condurre a pericolose devianze sottoforma di spettacolarizzazione della pena robotica, nonché a pericolose regressioni a stadi giuridici arcaici in cui l'amministrazione della giustizia veniva indirizzata nei confronti di cose ed animali.

All'interrogativo *"what else is needed?"* è stata dunque fornita una risposta da più fronti e questa è stata *"a lot is (still) needed"*. Necessita, innanzitutto, una ulteriore evoluzione delle macchine alla stregua di quanto teorizzato dai sostenitori di una *IA forte*, ossia, una IA in grado di manifestarsi dal punto di vista fenomenico quale *res cogitans*: un essere pensante in grado di comportarsi *come se* fosse umano e, dunque, di simulare umana coscienza, comprensiva dei suoi aspetti empatici, emozionali, etici e pedagogici. Fino a quando il sistema AI non sarà in grado di interiorizzare la pena applicata alle sue singole unità robotiche e di permettere l'esplicazione della reale funzione di questa, dunque, l'autonoma criminalizzazione di attanti robotici è destinata ad implodere nell'inutilità di una sanzione penale applicata ad una macchina e priva – al contrario di quella applicata agli umani e ai loro enti e società – di apprezzabili riflessi umani, sociali e culturali.

Ad ogni modo, anche le più convincenti critiche all'autonoma e diretta responsabilità penale delle macchine non hanno saputo disegnare un modello di imputazione alternativo in in grado di colmare il vuoto di tutela penale che aleggia intorno ai fatti commessi dalle macchine.

In sostanza: chi risponderà – se non le macchine – dei fatti di reato commessi dalle macchine?

I.2. LE POSIZIONI DI GARANZIA: L'UTILIZZATORE, IL PRODUTTORE, IL PROGRAMMATORE.

Elusa la possibilità di sottoporre le macchine ad autonoma responsabilità penale, non resta che da individuare una serie di figure umane che possano fungere da *"garanti"* dei fatti commessi dal sistema AI; figure che – pure estranee alla commissione attiva del fatto di reato – ne rispondono in ragione del combinato disposto dell'art. 40 cpv. c.p. *("non impedire un evento che si ha l'obbligo giuridico di impedire equivale a cagionarlo"*) e delle singole norme che nel futuro ordinamento postuleranno posizioni di garanzia in materia AI, ossia, delle singole norme che nel futuro ordinamento postuleranno in capo agli umani *"l'obbligo giuridico"* di impedire la commissione dei fatti di reato dell'AI.

In questo senso, la possibile allocazione di responsabilità si muove in una triplice direzione: in capo all'utilizzatore di applicativi AI; in capo al produttore di applicativi AI; in capo al programmatore di applicativi AI.

L'allocazione di responsabilità in queste direzioni può avere una logica cogente dal punto di vista della responsabilità civile, in seno alla quale il criterio di responsabilità oggettiva si configura quale valido criterio di attribuzione di eventi e

rispettivi danni economici. In particolar modo, ottimale si configura, sotto il versante dell'analisi economica del diritto, l'allocazione in capo al produttore di sistemi AI di responsabilità oggettiva per i danni da questi cagionati: non solo perché il produttore – ricavando reddito dalla commercializzazione di sistemi AI – sarà il soggetto meglio posizionato per sopportare il peso economico delle perdite da questi generate, ma altresì perché il costo patrimoniale di tali danni si collocherebbe in capo al produttore soltanto in via temporanea, per poi essere disperso tra la clientela tramite politiche di prezzo incorporanti i costi complessivi dell'attività di impresa.

Così non è per la responsabilità penale, dove il principio di colpevolezza sancito dall'art. 27 Cost. impone – relativamente alla responsabilità omissiva impropria *ex* art. 40 cpv c.p. – l'esistenza di una condotta impeditiva dell'evento penalmente rilevante; una condotta impeditiva materialmente implementabile dall'agente umano al fine di scongiurare il verificarsi dell'evento di reato.

Assegnare una posizione di garanzia *ex* art. 40 cpv c.p. in capo all'utilizzatore, produttore o programmatore di sistemi AI, dunque, significa postulare una costante possibilità di controllo e monitoraggio in capo all'essere umano in relazione al funzionamento di sistemi AI; nello specifico, significa postulare la possibilità dell'essere umano di comprendere e monitorare costantemente il processo computazionale dell'AI, così da intervenire in caso quest'ultimo stesse strutturando un processo di *decision making* orientato alla commissione di fatti di reato.

Eppure, così non è. La possibilità di comprensione e di intervento impeditivo da parte dell'essere umano – ivi compresa quella del programmatore, produttore o utilizzatore di AI – sembra infatti annichilirsi nell'opacità dei *black box algorithms*, nell'impossibilità umana di conoscere e governare il processo deliberativo delle macchine.

Una opacità che tra l'altro si disperde ulteriormente nel labirinto del *cloud computing*, ossia, nelle strutture informatiche remote – esterne alla fisicità dei singoli *hardware* – che guideranno il processo cognitivo e deliberativo delle macchine. In sostanza, ogni singola macchina AI sarà deputata, al più, alla sola acquisizione di dati e informazioni dalla realtà ad essa circostante. Questi dati verranno in seguito trasmessi ad una struttura in *cloud* a cui attingeranno una vastità di applicativi AI; sarà dunque in remoto che i dati acquisiti dalle macchine verranno analizzati alla luce della mole di informazioni presenti in *cloud* e processati tramite algoritmo, così da restituire alla singola macchina l'*output* che la indurrà all'azione materiale.

È nel labirinto oscuro dei *black box algorhitms* e del *cloud computing* che sfuma dunque la possibilità di muovere a singoli agenti umani addebiti penali fondati su legittimi criteri di attribuzione di responsabilità improntati a colpevolezza. È qui che si evidenzia, una volta di più, il rischio di un vuoto di tutela rispetto ai fatti di reato commessi dalle macchine.

I.3. Oltre l'allocazione della responsabilità penale per i fatti commessi dalle macchine: l'AI illecita, l'AI come strumento del reato e l'AI come vittima del reato.

Sembra necessario far seguire al tema dell'allocazione della responsabilità penale dei fatti di reato commessi dalle macchine una valutazione utilitaristica del calcolo rischi-

benefici che l'introduzione dell'AI potrà apportare nell'ambito del tessuto economico-sociale del prossimo futuro.

È infatti indubbio che l'ordinamento giuridico dovrà attrarre nel perimetro di liceità soltanto quegli applicativi AI ritenuti, ad un'analisi preventiva, in grado di iniettare nella realtà quotidiana una serie di benefici economico-sociali maggiori rispetto ai rischi (eventualmente, anche di natura penale) a questi inscindibilmente connessi. È il caso, ad esempio, delle automobili a guida autonoma, dove plurimi studi di settore si sono già occupati di dimostrare come l'implementazione di siffatta tecnologia su larga scala sfocerà in una drastica riduzione del numero di incidenti stradali e, dunque, di lesioni personali e di morti a causa di questi. Ciò posto, il beneficio atteso in termini generali porterà la società (e, dunque, l'ordinamento giuridico) a tollerare la manifestazione di singoli rischi inscindibilmente connessi all'introduzione della nuova tecnologia, così come l'investimento di singoli pedoni o la causazione di singoli incidenti stradali generati da sporadiche disfunzioni dell'AI. Tali rischi verranno dunque attratti nell'area del ***c.d. "rischio consentito"*** e la disciplina giuridica dovrà occuparsi di governare il loro profilo tramite allocazione della relativa responsabilità, ivi inclusa quella di natura penale.

Non così, invece, per le possibili offese giuridiche connesse all'implementazione di tecnologie AI incorporanti aree di rischio esorbitanti rispetto alle finalità deputate all'azione della macchina ed ai benefici da queste attesi. Sotto questo versante, la direzione prescelta potrebbe dunque essere quella di limitare la tipologia di *task* assegnati all'AI e limitare così in via artificiale la sua autonomia nell'esecuzione dei compiti ad essa affidati. Può essere il caso degli applicativi AI utilizzati in ambito militare: pare arduo poter attrarre nell'area del *'rischio consentito'* l'eventualità dell'uccisione di civili innocenti sulla scorta del beneficio atteso dall'automazione AI del *modern warfare*. In questi casi, dunque, l'implementazione AI potrebbe essere vietata *tout-court* o soggetta a stringenti limitazioni; ne consegue che il settore penale dell'ordinamento interverrebbe – non per trovare allocazione di responsabilità al rischio consentito e socialmente tollerato – bensì per reprimere l'eventuale introduzione nel sistema di rischi non consentiti, sottoponendo a sanzione penale i soggetti (umani) aventi implementato o utilizzato tecnologia AI in modo illecito poiché contrario a norme imperative di divieto.

In punto di perimetrazione del rischio consentito, dunque, la vera sfida giuridica sarà quella di irrorare l'ordinamento di divieti – eventualmente assistiti da sanzione penale – volti a limitare la potenzialità AI solo ed esclusivamente nei settori caratterizzati da rischi realmente intollerabili e non gestibili, così da non porre artificiosi ed irragionevoli freni giuridici alle potenzialità dell'innovazione tecnologica e sacrificare sull'altare di una logica precauzionale eccessivamente generalizzata i benefici insiti nella concreta applicazione delle macchine alla società contemporanea.

Problemi giuridici di matrice tradizionale vengono altresì evidenziati dall'eventualità che postula l'essere umano abusare dello strumento AI ai fini di asservirlo ad un suo intento criminale pre-ordinato, così come dall'ipotesi che vede l'essere umano servirsi dello strumento AI in maniera imprudente, negligente, imperita o contraria a specifiche norme precauzionali.

In entrambi i predetti casi, il centro di imputazione di responsabilità penale sarà indubbiamente umano, poiché l'AI si configurerà nelle mere vesti di oggetto (non soggetto) di diritto; nello specifico, con particolare riguardo ai

reati dolosi, nelle mere vesti di uno ***'strumento del reato'***, al pari del coltello usato dall'omicida o della bomba di cui si serve il terrorista. Al più – anche a voler riconoscere all'AI una soggettività giuridica più o meno estesa – essa assumerebbe in siffatti casi il ruolo di ***'autore mediato'*** del reato, ossia, il ruolo di una soggettività giuridica coinvolta nella realizzazione materiale della fattispecie criminosa eppure estranea alla reazione punitiva dell'ordinamento, poiché strumentalizzata alla commissione del reato sulla scorta di un errore indotto dall'artifizio posto in essere da un soggetto terzo.

Ferma restando l'univocità del centro di imputazione di responsabilità penale in capo all'essere umano, esempi di questo tipo inducono ad una riflessione circa l'idoneità delle fattispecie di reato attualmente previste dall'ordinamento a reprimere le potenzialità criminali scaturenti dal futuro abuso dell'AI, soprattutto alla luce della sua diffusione a larga scala nel circuito economico e sociale, pubblico e privato.

Non è infatti da escludersi – e, anzi, si ritiene probabile – che la consumazione di fattispecie di reato (sia pure tradizionali) mediante lo strumento AI riveli una attitudine lesiva tale da giustificare una risposta dell'ordinamento che passi per il tramite di una disciplina giuridica *ad hoc*, sia in veste di nuove fattispecie incriminatrici, sia in veste di introduzione di nuovi elementi accessori (circostanze aggravanti, *in primis*) alle fattispecie incriminatrici attualmente esistenti.

Un'ultima riflessione, infine, merita la possibilità di riconoscere all'AI il possibile ruolo di **vittima del reato** commesso dagli esseri umani. Esempi degni di nota provengono dal futuro mercato dei *sex toys*, ove è facile prevedere lo sviluppo di *robot* – non solo dotati di organi genitali artificiali – ma altresì dell'intelligenza artificiale necessaria ad interagire in modo evoluto con l'essere umano utilizzatore. E non pare potersi escludere la produzione di *sex toys* con le sembianze di minorenni in giovane età. Necessario chiedersi, fin da ora, quale sarebbe, in tali casi, la migliore reazione improntabile dal settore penale dell'ordinamento: l'inerzia, sulla scorta della ravvisata assenza di una effettiva lesione alla sfera sessuale umana o sulla scorta del possibile valore preventivo di macchine AI che, sfogando il bieco istinto carnale, neutralizzano le pulsioni sessuali nei confronti dei minorenni umani; la criminalizzazione del fenomeno, sulla scorta della ravvisata lesione di un comune (e invero a tratti retorico e paternalistico) sentimento del pudore e del buon costume presente in società; l'introduzione di fattispecie di reato a tutela delle macchine AI in quanto tali – vere e proprie persone offese del reato – oppure a tutela del sentimento empatico che l'essere umano nutrirà nei confronti di quest'ultime, così come già avviene, d'altronde, a tutela degli animali.

II. A.I., Procedimento Penale e Law-Enforcement

Il procedimento penale vive di forma.

Il diritto formale – ossia, la disciplina giuridica di una pluralità consequenziale e logicamente connessa di atti che conduce dall'emersione di una notizia di reato all'emissione di una sentenza definitiva – costituisce primaria garanzia dei diritti costituzionali riconosciuti all'indagato-imputato.

È il rispetto del diritto formale a garantire omogeneità di celebrazione tra diversi procedimenti penali, garantendo così il rispetto del principio di uguaglianza *ex* art. 3 Cost. È il rispetto del diritto formale a condurre alla formazione di provvedimenti giurisdizionali

trasparenti, inglobanti decisioni accessibili, controllabili e, come tali, censurabili dal loro destinatario. È il rispetto del diritto formale, in ultima istanza, a garantire la *"fairness"* del processo penale ai sensi del combinato disposto dell'art. 111 Cost e dell'art. 6 CEDU; un processo penale normato *ex lege*, celebrato in regime di parità delle parti dinnanzi ad un Giudice terzo ed imparziale, nell'ambito del quale l'imputato vanta il diritto di confrontarsi in contraddittorio con il proprio accusatore in merito agli elementi di prova addotti a suo carico.

È proprio la rigidità di forma del processo penale (o meglio, l'abuso retorico – e a tratti opportunistico – di questa rigidità), tuttavia, a condurre sempre più spesso il processo penale verso una vetustà ostinata ed irragionevole; una patologia, quest'ultima, il cui primo sintomo si ravvisa nella costante e reiterata incapacità del processo penale di assorbire qualsiasi innovazione tecnologica proveniente dal mondo esterno.

In questo senso, la pandemia da Covid-19 ha fornito brillanti esempi.

È servita una pandemia su scala mondiale per tollerare – in via del tutto eccezionale – la possibilità per le parti di procedere al deposito atti via PEC, in deroga alla imposizione normativa del deposito cartaceo, ad oggi costituente la regola nell'ambito del procedimento penale. Non solo: l'assenza di una regolamentazione normativa del deposito a mezzo PEC – oltre a porre in dubbio la validità dei depositi telematici effettuati in pieno *lockdown* – ha imposto un immediato ritorno al cartaceo nel periodo di latenza tra la fine della prima ondata e l'ascesa della seconda ondata epidemiologica. Solo l'avvento della seconda fase dell'epidemia – ormai giunto – sta inducendo il legislatore a disciplinare in via normativa il deposito telematico di atti (si veda il *d.l.* ristori di prossima emanazione); ciò nell'auspicio che tale innovazione possa superare il suo carattere di misura temporanea e tramutarsi in una soluzione strutturale che introduca nel procedimento penale *assets* di lungo periodo quali, ad esempio, la completa digitalizzazione dei fascicoli.

E ancora: è il caso della c.*d. udienza telematica*, ossia, della celebrazione dell'udienza penale a distanza, in un luogo diverso dal Tribunale e per il tramite di sistemi di video-conferenza. La soluzione – la cui ipotizzazione e temporanea implementazione è stata resa possibile solo ed esclusivamente dal culmine della prima ondata pandemica – ha trovato attriti e resistenze di grandissimo momento: ciò – non solo, come è ovvio – in relazione alla fase istruttoria del processo penale, laddove il principio di oralità e immediatezza che governa l'assunzione della prova nell'ambito del giusto processo preclude l'utilizzo di sistemi di video-conferenza, bensì anche nell'ambito di ulteriori fasi del procedimento, in relazione alle quali nessun ostacolo teorico si sarebbe seriamente opposto alla loro implementazione. E ancora: l'immediato accantonamento della soluzione tecnologica all'esito della prima ondata epidemica ha precluso agli operatori del diritto – tendenzialmente, non formati all'utilizzo dello strumento – l'acquisizione del metodo necessario ad un governo efficiente dell'udienza telematica; circostanza, quest'ultima, che farà sentire ora il suo peso, alla vigilia della re-introduzione dello strumento per mano del già citato *d.l.* ristori.

In ultima analisi, la refrattarietà del procedimento penale all'evoluzione tecnologica – a tratti giustificata da cogenti esigenze di giustizia sostanziale e tutela delle posizioni giuridiche dei singoli – esula troppo spesso dal

fondamento razionale di queste esenzioni, sfociando così nel patologico.

Questo, dunque, il fragile e acerbo terreno su cui nel prossimo futuro potrebbe innestarsi un fenomeno dirompente ed inedito quale a tutti gli effetti è l'implementazione dell'AI.

Dinnanzi ad un fenomeno – quello dell'AI – che secondo ormai unanime previsione a distanza di qualche anno *"sarà ovunque"*, non è certo che il processo penale potrà permettersi una chiusura ed un rigetto totale, sancendo così una forte cesura tra la sua realtà e la realtà sociale esterna, che peraltro è chiamato a giudicare. Non è certo, tuttavia, nemmeno che il processo penale – soprattutto partendo da uno stadio di deprivazione tecnologica – possa assimilare in modo efficiente, improntato a giustizia e costituzionalmente orientato un fenomeno complesso come quello AI.

II.1 I PROFILI DI INTERFERENZA TRA AI, PROCEDIMENTO PENALE E *LAW-ENFORCEMENT*.

L'esperienza nazionale e internazionale dello scorso decennio e le possibili prospettive future ci consegnano tre principali filoni di indagine in merito alla potenziale implementazione AI nell'ambito del procedimento penale:

- A.I. quale ausilio della polizia amministrativa in fase di *law-enforcement*. Trattasi di un momento antecedente – ma prodromico – all'instaurazione del procedimento penale, nell'ambito del quale l'Autorità agisce in una prospettiva *ante delictum* ai fini della prevenzione generale di fatti illeciti penalmente rilevanti. Il fenomeno è noto sotto la locuzione di ***"predictive policing"***;

- A.I. quale ausilio dell'organo giudicante nella valutazione della pericolosità sociale dell'imputato. Trattasi di una valutazione – pur diversa da quella inerente al binomio innocenza/colpevolezza – ma in grado di incidere sotto molteplici profili sulla posizione giuridica dell'indagato-imputato-condannato. Una valutazione dunque centrale nell'ambito del processo penale e che potrebbe essere compiuta o approfondita per il tramite dell'utilizzo di algoritmi predittivi noti come "***risk assessment tools***";

- A.I. quale ausilio dell'organo giudicante nella valutazione della colpevolezza dell'imputato. La soluzione tecnologica implementabile al fine è nota come ***"automated decision system"*** e sottende la valutazione AI del compendio probatorio del procedimento penale, finalizzata alla produzione di un *output* circa l'innocenza o colpevolezza dell'imputato. *Output* AI da porre ad ausilio – o, secondo le visioni più pioneristiche, in sostituzione – della decisione umana del giudice.

II.1.A *PREDICTIVE POLICING*.

Il concetto di fondo è quello narrato dal mai dimenticato Philip K. Dick nel romanzo *"The Minority Report"*: la computazione AI di una potenzialmente sconfinata mole di dati è in grado di individuare, tra questi, delle connessioni significative e difficilmente percepibili dall'intelligenza umana al fine di perimetrare tanto le aree geografiche a maggior rischio criminale (*hotspots*) quanto l'eventualità, la modalità e la tempistica di commissione di futuri reati da parte di determinati soggetti (*crime linking and profiling*).

Nel 1956 Philip K. Dick ne ha trattato in uno scritto di fantascienza: oggi, di fatto, la polizia

predittiva è realtà concreta in ambito internazionale e nazionale.

È denominato *X-Law* il *software* originariamente predisposto dalla Questura di Napoli ed in grado – tramite computazione AI delle informazioni contenute nelle denunce sporte dai cittadini – di geolocalizzare le aree connotate da maggior rischio criminale prospettico, consentendo così una mirata ed efficiente predisposizione delle forze dell'ordine sul territorio. La soluzione domestica deriva a sua volta da esperimenti condotti su base internazionale e, ad oggi, divenuti strumenti utilizzati su larga scala. Già da anni, infatti, USA e UK si avvalgono di *Predpol*, un *software* AI originariamente sviluppato dall'UCLA e oggi commercializzato da una società privata con base negli Stati Uniti. Le statistiche registrate dal *software* sono senz'altro significative: si evidenzia una riduzione del tasso di criminalità che spazia dal 20% al 70%, a seconda del titolo di reato preso in considerazione (*in primis*, reati venali come il furto o la rapina).

È invece denominato *Keycrime* il *software* elaborato dalla Questura di Milano capace di incrociare dati provenienti da una molteplicità di fonti – info contenute in *data base* delle Autorità o ricavate dagli impianti di video-sorveglianza*, ,* info presenti nel *web* e relative a reati analoghi, *etc* – al fine di profilare l'attitudine criminale dei delinquenti seriali e prevedere così modalità e tempistiche della commissione di futuri fatti di reato. E la soluzione domestica non è un *unicum* nello scenario internazionale: strumenti assimilabili trovano corrente utilizzo in Germania, in Inghilterra e negli Stati Uniti. Il funzionamento di tali strumenti poggia sull'esistenza di *pattern* criminali; *pattern* che l'AI è in grado di individuare ed elaborare, ottenendo così una fotografia talmente lucida di quanto successo in passato da poter essere posta a valida previsione prognostica circa il comportamento futuro di soggetti appartenenti a determinati *cluster* criminali.

Capacità predittive del tipo di quelle ora analizzate inducono a perimetrare singole aree geografiche ad elevato rischio criminale, sia generico (gli *hotspots*), sia inerente a determinati individui o categorie di individui (*crime linking and profiling*). Unitamente al rischio criminale, dunque, ad essere perimetrato è altresì il rischio insito in determinate aree per la incolumità della popolazione, ivi compresa quella degli agenti delle forze dell'ordine chiamati ad intervenire nelle predette zone proprio in virtù della potenzialità criminogena ivi individuata. Incolumità che, tuttavia, potrebbe essere tutelata mediante ulteriori applicativi AI: macchine robotiche – non necessariamente umanoidi, si pensi ai droni – in grado di utilizzare le informazioni raccolte dai propri sensori ambientali al fine di portare a termine una serie di attività di polizia quali il pattugliamento, la sorveglianza e la neutralizzazione di atteggiamenti sospetti o criminali. Dal romanzo di Philip K. Dick al Robocop cinematografico di Paul Verhoeven il passo è dunque breve; una visione fantascientifica di fine anni '80 è, trent'anni più tardi, alle soglie della realtà: non mancano le giurisdizioni statali i cui corpi di polizia vantano in dotazione agenti robotici equipaggiati con avanzati sistemi di riconoscimento facciale e con armi stordenti o letali, attivate dall'esito della computazione AI dei dati acquisiti dalla realtà esterna per il tramite del processo di autoapprendimento noto come *machine learning*.

II.1.B. *RISK ASSESSMENT TOOLS*.

Gli applicativi AI, come visto, paiono in grado di giungere ad accurate stime circa il rischio

criminogeno generalmente insito in determinate aree. Se così è, valutazioni ancor più accurate ed approfondite potranno essere condotte dall'AI in merito alla pericolosità sociale di singoli individui pre-determinati e, in particolar modo, in merito al pericolo di commissione di futuri reati da parte di soggetti già sottoposti a procedimento penale.

Gli applicativi AI impiegati al fine sono noti come *risk assessment tools* e rappresentano il ponte d'oro che conduce l'AI dalla funzione di ausilio alla polizia amministrativa in ottica general-preventiva alla funzione di ausilio ai Tribunali in ottica special-preventiva.

Non vi sono (ad oggi) esempi nazionali da sottoporre ad analisi. Lo sguardo si rivolge dunque all'esperienza statunitense, dove tali applicativi AI – PSA e COMPASS i più utilizzati – intervengono in diverse fasi del procedimento penale, condizionando così su più livelli il grado di limitazione della libertà personale inflitto all'imputato:

- in **fase cautelare** (*pre-sentecing*), ove gli applicativi AI intervengono per valutare il rischio che l'indagato-imputato si sottragga al futuro procedimento penale e, quindi, per determinare l'opportunità di applicazione dell'istituto della cauzione (*bail*), il cui pagamento consente la cessazione del regime di carcerazione preventiva in attesa del processo;

- in **fase decisionale** (*sentecing*), ove gli applicativi AI intervengono al fine di adiuvare il Giudice nella commisurazione della pena da infliggere all'imputato colpevole e, in particolar modo, nell'individuare il *quantum* di pena necessario a disinnescare, in ottica special-preventiva, la pericolosità sociale dello stesso, fungendo da deterrente per la commissione di futuri reati;

- in **fase esecutiva,** ove gli applicativi AI intervengono in ausilio alla valutazione del Giudice circa l'opportunità di concedere al condannato tanto la liberazione condizionale (*parole*) quanto la possibilità di scontare la pena tramite misura alternative alla detenzione (*probation),* parametrando tali valutazioni al rischio di recidivanza del condannato in regime di scarcerazione.

In linea teorica, i margini per l'implementazione di *risk assessment tools* nell'ambito del procedimento penale italiano non mancano e, anzi, brillano per varietà ed estensione:

- valutazione del *periculum* di recidivanza (soprattutto specifica) ai fini dell'applicazione di misure cautelari coercitive o interdittive *ex* art. 274 *lett. c)* c.p.p.;
- valutazione della pericolosità sociale dell'imputato ai fini dell'applicazione di misure di sicurezza *ex* art. 202 c.p.;
- valutazione del *quantum* di pena da applicare in concreto in ragione della capacità a delinquere del reo *ex* art. 133 c. 2 c.p.;
- valutazione del rischio di recidivanza ai fini della concessione del beneficio della sospensione condizionale della pena *ex* art. 164 c.p.;
- valutazione della pericolosità sociale degli individui ai fini dell'applicazione di misure di prevenzione *ex* d.lgs. n. 159/2011.

Una valutazione centrale nel procedimento penale come quella inerente alla pericolosità sociale e al rischio di recidivanza dell'indagato potrebbe dunque essere rimessa – anziché al solo giudizio umano, guidato dalle regole codicistiche – altresì alla computazione di dati da parte dell'AI nell'ambito di un modello attuariale-statistico noto come *evidence-based approach.*

Diversi i fattori (*evidence*) potenzialmente computabili dall'AI ai fini di tale valutazione: età,

sesso, origine etnica, livello di scolarizzazione, situazione familiare e lavorativa, posizione sociale, precedenti penali, precedenti esperienze carcerarie, luoghi e persone frequentate, presenza di rei nella cerchia familiare o sociale di riferimento, luogo di residenza, di controllo degli impulsi, precedenti ospedalieri, consumo di droghe, psicopatie, *etc*.

II.1.c. *AUTOMATED DECISION SYSTEMS.*

I *software* AI sono ad oggi perfettamente in grado di esaminare, in tempi rapidissimi, il dato normativo e giurisprudenziale, nonché di estrarne la conoscenza necessaria alla decisione di specifici casi sottoposti alla loro attenzione; di più, i *software* AI sono perfettamente in grado di applicare la *teoria dei giochi* al fine di strutturare strategie di negoziazione ritenute ottimali ai fini della risoluzione consensuale delle controversie tra le parti.

Anche in questo caso, nulla di più distante dalla fantascienza: l'utilizzo di algoritmi AI ad ausilio (o in sostituzione) del processo umano di *decision making* è già diventata la regola per molti soggetti operanti, anche in ambito europeo, nel settore finanziario, assicurativo, bancario e molti altri, tra cui il settore legale composto da avvocati e studi associati; non solo, nel settore pubblico della giustizia civile, commerciale ed amministrativa, la risoluzione *on-line* – demandata al *ruling* di algoritmi AI – di controversie di modico valore è alla vigilia della sua diffusione, così come dimostra, in ambito europeo, la recente iniziativa estone avviata in questo senso e, in ambito US, il profondo lavoro di ricerca e sintesi condotto dal Prof. Richard Susskind nel suo *"Online Courts and the future of Justice"*.

Stante la direzione in cui spinge l'inerzia del fenomeno, è spontaneo interrogarsi circa la possibilità che la macchina AI – dopo aver saturato il settore privato ed aggredito il settore pubblico della giustizia civile e amministrativa – approdi in seno al processo penale.
Fondamento razionale e orizzonte ideale di tale approdo sarebbe il raggiungimento di una giustizia penale matematica, scientificamente esatta, priva delle distorsioni che contaminano la fase cognitiva e la conseguente decisione umana; come tale, una giustizia penale equa ed omogenea, le cui decisioni – sempre più prevedibili con certezza – fungano da criterio di allineamento delle condotte dei singoli in fase preventiva.

La capacità della macchina – in sostituzione o, come è più auspicabile, ad ausilio normato del Giudice umano – di raggiungere l'obiettivo di una giustizia penale somministrata con rigore scientifico e con infinitesimale margine di errore è tutta da dimostrare: a rigor di logica, il *software* AI è in grado di portare a termine compiti ben più complessi, ma la sua solidità sul banco di prova della giustizia penale non è mai stata corroborata da alcun esperimento empirico.

Pertanto, diversi sono i fronti aperti in questo senso e, tra i principali, si annoverano i seguenti:

- la capacità della macchina di valutare le contraddizioni insite nelle dichiarazioni di un teste e apprezzarne, così, l'attendibilità, nonché la capacità della macchina di scorgere falsità e reticenze nelle dichiarazioni del teste senza ricorrere a test alterativi della sua libertà morale;
- la capacità della macchina di valutare l'idoneità prospettica degli elementi di accusa addotti dal Pubblico Ministero in udienza preliminare ai fini dell'emissione di una sentenza di condanna all'esito del dibattimento e, dunque, ai fini della valutazione dell'opportunità di

celebrazione dell'intera istruttoria dibattimentale;

- la capacità della macchina di qualificare gli indizi alla stregua dei criteri di gravità, precisione e concordanza previsti dalla legislazione codicistica;
- la capacità della macchina di assimilare ed applicare il criterio del *'beyond any reasonable doubt'* ai fini del giudizio circa la colpevolezza dell'imputato, alla luce della linfa, non solo statistica, ma altresì logica, che irrora lo standard probatorio imposto dal settore penale dell'ordinamento.

Il palcoscenico, insomma, è vasto e, per ora, deserto. Proprio su questo palcoscenico, tuttavia, potrebbe giocarsi il futuro del processo penale.

II.2. Il vuoto normativo dinnanzi all'innovazione tecnologica AI.

Il rapporto tra AI e processo penale è ad oggi selvaggio ed indomito: si sono finora registrate scarse occasioni di approfondito confronto tra la scienza giuridica e quella cibernetica. Un confronto, tuttavia, quanto mai essenziale per l'avvento di una regolamentazione puntuale ed efficiente che possa assistere, a tempo debito, la fattiva introduzione dell'AI nell'ambito del processo penale.

L'embrione di tale futura regolamentazione è stato impiantato in seno all'UE il 4 dicembre 2018 per il tramite dell'adozione della *"Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti"* da parte della Commissione per l'efficacia della giustizia (CEPEJ). La *Carta etica* ha rappresentato uno snodo fondamentale nel percorso regolamentare prospettico: *"preso atto della crescente importanza dell'intelligenza artificiale nelle nostre moderne società e dei benefici attesi quando questa sarà pienamente utilizzata al servizio dell'efficienza e qualità della giustizia"*, la Carta ha infatti voluto individuare alcune fondamentali linee guida alle quali *"dovranno attenersi i soggetti pubblici e privati responsabili del progetto e sviluppo degli strumenti e dei servizi della IA"*.

In ultima analisi, la *Carta* ha voluto enucleare i principi che fungeranno da *grundnorm* per il futuro apparato regolatore del fenomeno AI; trattasi dei seguenti – a parere di chi scrive, ben individuati – principi:

- rispetto dei diritti fondamentali;
- non discriminazione;
- qualità e sicurezza;
- trasparenza, imparzialità e correttezza;
- garanzia del controllo umano.

La Carta Etica apre dunque alla sfida del futuro: strutturare una regolamentazione giuridica – teleologicamente orientata ai predetti principi fondamentali – finalizzata alla massimizzazione dei benefici derivanti dall'applicazione dell'AI e alla contestuale minimizzazione gli effetti collaterali (se non nocivi) che l'introduzione di ogni innovazione tecnologica rischia di portare con sé.

Trattasi di una sfida complessa. Le potenzialità degli applicativi AI sono evidenti, ma altrettanto significative sono le relative criticità.

Gli applicativi AI vengono correntemente usati in funzione di ***polizia predittiva*** nell'assenza di ogni regolamentazione – europea o nazionale – in materia. Nessuna risposta regolamentare viene dunque fornita all'interrogativo in materia *privacy* connesso allo *storage* ed all'utilizzo di dati personali, potenzialmente sensibili, raccolti dall'agente AI per il tramite di sensori iper-evoluti

quali quelli capaci di riconoscimento facciale tramite parametro biometrico; dati che peraltro – al pari di ogni substrato digitale dell'era post-moderna – rischiano di essere oggetto di sottrazione ed utilizzo indebito nella logica *cybercrime*. Allo stesso modo, nessuna garanzia regolamentare viene fornita in merito alla qualità degli *input* che alimentano la computazione dell'AI agente in funzione di polizia predittiva, così come in merito alla completezza e coerenza dei *database* informativi da cui tali *input* sono tratti; *input* che, tuttavia, condizionano in via determinante l'azione dell'AI e, quindi, influenzano la possibilità di commissione da errori da parte di questa, ivi compresa la possibilità di sottoporre ad ingiustificata privazione la libertà personale dei cittadini. Da ultimo, la regolamentazione non interviene al fine di apporre eventuali correttivi umani alla '*militarizzazione'* di singole aree qualificate come *hotspot* dall'AI con conseguente decremento della sorveglianza di altre zone ritenute a minor rischio criminale; zone che, tuttavia, proprio per questa ragione possono essere prese di mira dalla delinquenza, che potrebbe così creare nuovi ed improvvisi *hotspot* non immediatamente individuabili dall'AI nelle more della formazione di dati strutturati al riguardo. Evidente, alla luce di tutto ciò, l'opportunità di una regolamentazione che cali nei singoli tratti del fenomeno i principi di **tutela dei diritti fondamentali**, **qualità, sicurezza** e **garanzia di controllo umano** sanciti dall'Unione Europea.

Diverse, invece, le criticità di cui la regolamentazione dovrà farsi carico in relazione agli algoritmi AI usati all'interno del processo penale in funzione di ausilio alla valutazione sia della pericolosità sociale ***(risk assessment tools)*** che della colpevolezza dell'indagato-imputato ***(automated decision system)***. Se la direzione prescelta sarà quella di dotare l'organo giudicante di uno strumento AI sulla scorta del modello USA – ferma restando la sicurezza e qualità del servizio – il punto focale verrà a condensarsi infatti sotto i profili della **trasparenza** e della **non discriminazione** della computazione AI asservita al fine.

Il **principio di trasparenza**, in particolare, crea un primo attrito con i diritti di proprietà intellettuale che potrebbero insistere sul *software* AI utilizzato dall'organo giudicante in ausilio alle proprie valutazioni. È esattamente ciò che succede quotidianamente negli USA tramite utilizzo del già citato *software* COMPASS, *software* coperto da segreto industriale che impedisce la divulgazione di informazioni relative al suo processo di funzionamento. La circostanza è stata sottoposta al giudizio della Corte Suprema del Winsconsin nel celebre ***caso Loomis***, originato dal ricorso proposto dall' omonimo imputato, la cui pena detentiva era stata commisurata con l'ausilio delle valutazioni compiute dal *software*, delle quali era noto l'esito ma imperscrutabile il processo di valutazione che a tale esito aveva condotto. Il *ruling* della Corte Suprema, tuttavia, ha finito con il convalidare il legittimo utilizzo dell'*output* prodotto dal software da parte della Corte territoriale; ciò sulla scorta della considerazione per cui tale *output* era stato utilizzato in veste di mero fattore ausiliario alle valutazioni condotte dal Giudice umano, il quale aveva comunque steso un giudizio *'umano'* e razionalmente motivato in punto di quantificazione della pena inflitta al ricorrente. Nell'alveo della *case-law* USA pare dunque iniziare a delinearsi una prima bozza di gerarchia tra i principi cardine che sorreggono l'utilizzo dell'AI nell'ambito del processo penale; una gerarchia per cui il principio di trasparenza – la cui esigenza di rispetto resta comunque riconosciuta – pare soccombere dinnanzi alla tutela posta dal principio di garanzia del controllo umano

sull'esito finale delle valutazioni condotte dalla macchina.

Il principio di trasparenza, tuttavia, trova un secondo e più significativo ostacolo nel vuoto cognitivo che si raccoglie in seno al fenomeno noto come ***"algorithm black box"***: trattasi, in sostanza, della riconosciuta incapacità umana di prevedere *ex ante* e ricostruire *ex post* i processi computazionali che consentono all'AI di produrre determinati *output* a partire dall'analisi degli *input* introdotti sotto forma di dati. Caratteristica fondante e distintiva dell'AI, d'altro canto, è proprio la sua capacità di strutturare ed implementare processi decisionali autonomi – non previamente istruiti nella macchina dall'essere umano – sulla scorta di quello che è noto come modello di *machine learning* o *deep learning*. Trattasi, dunque, di una frizione di grande momento con il principio giuridico di trasparenza, poiché insiste sul profilo tecnologico che costituisce il cardine del funzionamento dell'intero sistema AI: la macchina è definita intelligente perché in grado di sviluppare un *pensiero* autonomo a partire dai dati di *input*; un pensiero cangiante e in costante perfezionamento, in grado di evolvere al punto che lo stesso designer dell'algoritmo di base può giungere a non essere in grado di spiegare e giustificare compiutamente gli *output* restituiti dalla macchina programmata con lo stesso algoritmo da lui disegnato.

Ne consegue che – se la volontà collettiva sarà quella di convalidare le decisioni assunte dalle macchine AI sulla scorta della loro (eventuale) comprovata validità empirica – la deriva possibile sarà quella di sviluppare un tessuto economico-sociale fondato sulla *'fede'* nelle macchine, delle vere e proprie ***"trust machines"*** a cui verrà delegata una serie rilevante di compiti, istruiti nell'imperscrutabile oscurità della *black box* dell'algoritmo da una intelligenza diversa da quella umana eppure riconosciuta come meglio posizionata al fine. Il nodo è denso di criticità, sul versante sociale, economico, scientifico, antropologico e psicologico. Sul versante giuridico – e, in particolare, penalistico, che qui interessa – si evidenzia come un siffatto *'salto della fede'* nei confronti delle macchine, se non adeguatamente indirizzato, potrebbe svuotare dall'interno il contenuto del giusto processo costituzionalmente tutelato.

Il rischio sarebbe quello di un processo penale in cui la prova scientifica restituita dall'AI – proprio perché imperscrutabile nelle sue logiche – non potrà essere fatta oggetto di valutazione peritale e, soprattutto, di falsificazione all'esito del contraddittorio tra le parti, con evidente *vulnus* al diritto di difesa dell'imputato.

Il rischio sarebbe inoltre quello di un processo penale restituente provvedimenti giurisdizionali parzialmente sguarniti di motivazione razionale, poiché parzialmente fondati su valutazioni condotte da macchine con logiche non intellegibili dalla mente umana; ciò in lesione, non solo del diritto di difesa dell'imputato, ma della sua facoltà di contestare e criticare le prove a suo carico, vero nucleo fondante della nozione di giusto processo.

E ancora, il rischio sarebbe quello di consentire all'algoritmo – ad insaputa di tutti i soggetti del processo penale – di elaborare decisioni fondate sull'analisi di una serie di fattori (tendenze sessuali, preferenze commerciali di consumo, pensiero pro-criminale, *etc.*) che l'imputato – in esercizio del suo diritto al silenzio costituzionalmente tutelato – avrebbe facoltà di escludere dall'ambito di cognizione del processo penale. Una serie di fattori, quest'ultimi, che rischierebbero inoltre di **smantellare il paradigma del *'diritto penale del fatto'*** – imperante fin dalla teorizzazione del diritto

penale moderno brillantemente evidenziata nel saggio di Beccaria *"Dei Delitti e delle Pene*" – fino a ricondurlo in seno a teorizzazioni abbandonate da tempo quali quelle di matrice positivista e antropologico-criminale elaborate da Cesare Lombroso, il cui *focus* è posto non sul fatto penalmente rilevante commesso dal reo, bensì sulla sua attitudine delinquenziale, così come desunta dalle sue caratteristiche biologiche e comportamentali. Una serie di fattori, sempre quelli più sopra ricordati, tra l'altro potenzialmente non eleggibili a criteri fondanti un provvedimento giurisdizionale, poiché contrari ad un principio cardine dell'ordinamento democratico-pluralista come quello di non discriminazione.

In ultima analisi, le criticità sono molteplici e variegate. Non si è pronti a delegare ad un codice algoritmico il discernimento tra ciò che è penalmente rilevante e ciò che non lo è, così come non si è pronti ad affidare ad un algoritmo la decisione circa l'innocenza o la colpevolezza dell'imputato, soprattutto per il tramite di un processo che rischia – proprio in ragione dell'introduzione dello strumento AI – di perdere alcuni tratti del suo carattere di *'fairness'*. Ad oggi, in materia penale, *"code is (not) law"* e forse non lo sarà nemmeno nel prossimo futuro. Allo stesso tempo, tuttavia, le potenzialità derivanti dall'introduzione di applicativi AI in seno al processo penale sono evidenti e sterminate: impossibile pensare di imploderle nell'aprioristico rifiuto anziché esploderle in un serrato confronto tra la scienza giuridica e quella cibernetica.

CONTATTI

Giuseppe Fornari
Founding Partner

g.fornari@fornarieassociati.com

Nicolò Biligotti
Associate

n.biligotti@fornarieassociati.com

Martina Cupella
Associate

m.cupella@fornarieassociati.com

E: milano@fornarieassociati.com

E: roma@fornarieassociati.com

Fornari e Associati Studio Legale

www.fornarieassociati.com