GIURISPRUDENZA PENALE

Temere l'intelligenza artificiale? Il ruolo del giudice e l'insostituibilità della funzione autonoma di giudizio nel sistema penale.

di ***Nicola Scerbo***

Sommario. 1. Qualche cenno introduttivo. - **2.** La giustizia predittiva. - **3.** Esempi concreti d'impiego (il caso Loomies). - **4.** L'approccio europeo al tema (trasparenza e proporzione). - **5.** Il controllo umano (governance e due diligence). - **6.** *Law enforcment* ed attività investigative. - **7.** Il contesto penale. - **8.** Conclusioni

1. Qualche cenno introduttivo

Nell'introdurre il tema relativo all'impiego dell'Intelligenza artificiale nella giustizia penale ed i suoi potenziali rischi nel rispetto delle libertà fondamentali, è opportuno soffermarsi su alcuni aspetti teoretici e di ragionamento etico. Al riguardo, è interessante estrapolare alcuni frammenti di una recente intervista al prof. Luciano Floridi, docente di etica dell'informazione all'università di Oxford[1]. Alla domanda su cosa debba intendersi per A.I. lo stesso ha risposto che «si intende quasi sempre un software e non un hardware. È naturale per la pubblicistica ricorrere al robot per illustrare l'intelligenza artificiale. Ma la robotica è un angolo molto limitato della A.I. che abbiamo intorno a noi. Per rendersene conto basta ricordarsi dei suggerimenti che riceviamo su Netflix o su Amazon: quello è un pezzettino di A.I.». Dunque, sempre secondo il docente «si tratta di software che risolvono problemi con successo, in maniera efficiente. Al punto che, se dovessimo farlo noi – come si dice in gergo – "dovremmo essere intelligenti". Una logica controffatuale, che serve a spiegare che in realtà i software non sono intelligenti»[2]. Floridi cita l'esempio della lavastoviglie, che

[1] Per leggere l'intervista completa, si veda *Temere l'intelligenza artificiale? È come temere il demonio o le streghe…* di F. SPAESIANO del 20/12/2021 su quotidiano *il Dubbio*.

[2] L'esempio classico è quello degli scacchi: un banale gioco presente su *smartphone* vince sempre. Non perché sia più intelligente del giocatore, ma perché, nell'esecuzione del programma, ad egli è richiesta molta più intelligenza, di cui il *software*, invece, non ha bisogno. È più o meno come il corso dell'acqua in discesa da un pendio, essa trova sempre la strada più breve. Per tale motivo secondo la definizione comune, almeno dagli anni cinquanta, «l'*A.I.* è un qualsiasi sistema ingegnerizzato che risolve problemi con successo, in modo tale che se dovessimo

«ha successo perché costruisce un mondo intorno alle elementari capacità del robot che ha all'interno. Ecco cosa succede con la nostra A.I. Siamo noi a costruire un mondo intorno al software, che quindi si trova a "casa sua": quando Netflix ci suggerisce un film, è perché ha raccolto tantissime informazioni. Ed è tanto più bravo ad indovinare il prossimo film, quanto più usiamo la piattaforma. Senza dati da macinare l'A.I. non sa cosa dire». Di fatto i dati sono la chiave di tutto il sistema, senza gli stessi, l'intelligenza artificiale non avrebbe punti di riferimento e possibilità di autoapprendimento, sulla base di modelli di riferimento da cui trarre conclusioni e correlazioni, utili ad assumere la decisione più idonea al problema posto alla sua attenzione. Floridi stesso specifica che, «siccome i dati sono la linfa vitale della A.I., se abbiamo dati sporchi, parziali o incorretti, il sistema li userà e il risultato non potrà essere che negativo. Fa parte dei limiti del sistema: attinge dal passato, e se il passato è "storto", lo sarà anche il futuro. Così buona parte dei problemi che noi vediamo, in termini di ingiustizia e di bias ("pregiudizi" che un algoritmo tende a rafforzare), non sono del software, ma dei dati che abbiamo accumulato e sul quale il software è stato "esercitato". Il che denota come, se la società non andava bene da prima, non è certo colpa del software. La quale, semmai, automatizza e ingigantisce l'errore. Lo vediamo molto bene nei contesti ad alta sensibilità umana, da quello giudiziario, a quello medico, o nei servizi sociali e finanziari»[3]. Il principale elemento di

farlo noi dovremmo essere intelligenti».

[3] Per approfondimenti si veda la Rilevazione tecnologica "*L'Intelligenza Artificiale in banca: stato dell'arte e prospettive*" - Anno 2020 - Rapporto conclusivo pubblicata sul sito del CIPA. Come specificato sul sito della banca d' Italia la «"*Rilevazione sull' IT nel settore bancario italiano - Profili tecnologici e di sicurezza*", curata annualmente dalla CIPA (Convenzione Interbancaria per l'Automazione) in collaborazione con l'ABI, fornisce un quadro aggiornato sull'utilizzo dell'*Information and Communication Technology* nel settore bancario nazionale, con particolare riguardo alle scelte *IT* in materia di metodologie e tecnologie innovative a supporto dell'operatività bancaria». Allo stesso tempo il *GDPR* al punto 54 afferma che «Il trattamento di categorie particolari di dati personali può essere necessario per motivi di interesse pubblico nei settori della sanità pubblica, senza il consenso dell'interessato. Tale trattamento dovrebbe essere soggetto a misure appropriate e specifiche a tutela dei diritti e delle libertà delle persone fisiche. In tale contesto, la nozione di «sanità pubblica» dovrebbe essere interpretata secondo la definizione del regolamento (CE) n. 1338/2008 del Parlamento europeo e del Consiglio: tutti gli elementi relativi alla salute, ossia lo stato di salute, morbilità e disabilità incluse, i determinanti aventi un effetto su tale stato di salute, le necessità in materia di assistenza sanitaria, le risorse destinate all'assistenza sanitaria, la prestazione di assistenza sanitaria e l'accesso universale a essa, la spesa sanitaria e il relativo finanziamento e le cause di mortalità. Il trattamento dei dati relativi alla salute effettuato per motivi di interesse pubblico non dovrebbe comportare il trattamento dei dati personali per altre finalità da parte di terzi, quali datori di lavoro, compagnie di assicurazione e istituti di credito».

preoccupazione per chi si confronta con tali sistemi è il terrore che il costante e progressivo processo di umanizzazione della macchina possa portare ad una corrispondente sostituzione dell'essere umano nel suo ruolo di figura principale nell' esistenza naturale. Si tratta di una trasformazione che Floridi definisce quarta rivoluzione, essenzialmente di tipo epistemologico e di autocomprensione, della serie, «noi chi siamo? E perché siamo speciali? Speciali perché siamo al centro della festa e la festa è l'universo. Dopo che Freud ci ha detto che non siamo neanche al centro nel "mondo mentale", ci siamo trincerati dietro un'altra centralità: quella dell'informazione. Ma dagli anni '50 in poi, non siamo più neanche al centro dell'infosfera. Ecco la "quarta rivoluzione", un cambio di paradigma che comporta un necessario ripensamento della nostra eccezionalità. E la nostra non è quella di chi gode della festa, ma di chi può organizzarla. Non possiamo stendere la coperta del ventesimo secolo. Dobbiamo scrivere un nuovo capitolo»[4].

2. La giustizia predittiva.

Il più interessante e, allo stesso tempo, inquietante impiego dell'intelligenza artificiale, soprattutto alla luce dei suoi possibili esiti sugli aspetti più delicati della libertà individuale, riguarda il cosiddetto impianto di giustizia predittiva, altrimenti inteso come sistema che consenta di prevedere il possibile esito di una controversia in virtù dei precedenti esiti di casi analoghi[5]. Come sostenuto dal presidente della Corte di appello di Brescia, Claudio Castelli, in un recente articolo apparso su "Questione giustizia", «In realtà la prevedibilità è un enorme valore e la predittività non è che lo sviluppo della prevedibilità delle decisioni. Prevedibilità significa certezza del diritto, perché quanto conta oltre alle norme sono le interpretazioni che le trasformano in diritto vivente, ovvero la concreta possibilità di godere di un diritto. Prevedibilità che si scontra con una legge spesso ambigua in cui la mediazione politica affida all'interprete il compito di sciogliere contrasti e con l'incerta tecnica legislativa. Ma anche con la discrezionalità lasciata all'interprete. Le diverse interpretazioni non dipendono solo e fondamentalmente da opzioni valoriali diverse, ma dalla complessità ed incertezza dell'attività di interprete in un

[4] *Temere l'intelligenza artificiale? È come temere il demonio o le streghe…* di F. SPAESIANO del 20/12/2021 su quotidiano *il Dubbio*.

[5] In materia il contributo ed il dibattito dottrinale sono enormi. Solo per citarne alcuni, CADOPPI, *Giudice Penale e giudice civile di fronte al precedente,* in *Indice penale,* 2014, p. 14 ss; COSTANTINO, *La prevedibilità della decisione tra uguaglianza e appartenenza*, Relazione all'XI assemblea degli osservatori civili, 2016; SALVANESCHI, *Diritto giurisprudenziale e prevedibilità delle decisioni: ossimoro o binomio*, Relazione all'XI assemblea degli osservatori civili, 2016; VIGANÒ, *Il principio di prevedibilità della decisione giudiziale in materia penale,* in *Diritto penale contemporaneo*, 19 dicembre 2016; XII Assemblea Nazionale degli Osservatori – Roma 2017 Gruppo di lavoro – Prevedibilità delle decisioni e dialogo fra i diversi gradi della giurisdizione.

momento ed in un assetto come l'attuale in cui abbiamo una realtà multi-fonte in rapida evoluzione che impongono la ricostruzione dell'ordinamento e della gerarchia delle fonti[6]». Le opzioni tecnologiche disponibili alimentano il timore circa la possibile visone meccanicistica del ruolo del giudice, edulcorato o, se si vuole, mascherato dall'impiego delle nuove tecniche nella prospettiva di decisione automatizzate ed indifferenti alla realtà umana[7]. Si tratta di una prospettiva molto inquietante, alla quale si potrebbe dare risposta concreta alimentando il grado di consapevolezza sull'esistenza di possibili fattori esterni ed interni, dal potenziale condizionante in merito a decisioni sensibili, nel tentativo di sterilizzarne gli effetti principali attraverso l'uso di una discrezionalità ragionata, motivata e trasparente. In pratica, non una vera e propria sostituzione del giudice con la macchina, ma un contributo che può essere erogato dall'intelligenza artificiale alla sua già innata capacità di elaborazione dei dati, ma velocizzandone notevolmente i tempi. Da tali considerazioni emerge come i sostenitori di un uso sostitutivo dell'A.I. mediante decisioni automatizzate, sull'assunto di possibili risvolti positivi in termini di oggettività, non suscettibilità a pressioni esterne, ma, soprattutto, tempistiche più brevi, non si rendano conto delle possibili conseguenze di una simile determinazione. Come già accennato, le decisioni delle varie A.I. derivano dalla qualità ed integrità dei dati impiegati nel processo elaborativo, sia negli aspetti processuali che nei precedenti giurisprudenziali; pertanto, elemento essenziale, nell'implementazione di tali sistemi, dovrebbe essere la costante ed agevole controllabilità dei parametri anzidetti. Come sostenuto anche dal giudice Castelli, «affidarsi ad una macchina inevitabilmente implica abbandonare il nostro sistema basato sulla possibilità di far rivedere e, eventualmente, correggere la decisione di primo grado con un giudice superiore di maggiore esperienza e collegialità. La macchina, una volta pronunciatasi, è di per sé inappellabile, un po' come la giuria nel processo accusatorio statunitense. Infine, dovrebbe farci pensare che l'elaborazione di una macchina deriva e parte dalle pronunce e dai precedenti di Tribunali e Corti composti da esseri umani. Migliaia, anzi milioni di precedenti, che vengono immagazzinati e rilavorati per trovare la soluzione al caso specifico. Precedenti che rispecchiano un'evoluzione degli orientamenti giurisprudenziali, ma che sono a loro volta espressione e portato dei cambiamenti sociali e di costume di una società e non solo delle modifiche normative. Affidarsi alle decisioni di una macchina significa fermare questa benefica osmosi tra realtà sociale, diritto e giurisprudenza ossificando le

[6] Per approfondimenti sul tema si veda *giustizia predittiva* a cura di C. CASTELLI, in Q*uestione Giustizia,* rivista *online, 08/02/2022.*

[7] Timori rappresentati anche dal *Report* del *Gruppo 1*, *Prevedibilità, predittività e umanità del giudicare* della XIII *Assemblea Nazionale degli Osservatori sulla giustizia civile.*

decisioni odierne e future all'oggi. Anche perché una volta che le decisioni sono automatizzate non c'è più il precedente umano, ma solo un precedente meccanizzato che non fa altro che ripetersi e confermarsi[8]». D'altro canto, alcuni elementi di riflessione su eventuali pregi meritano un breve cenno. In generale, ciò che maggiormente caratterizza il processo decisionale umano non è solo riferibile ad elementi di valutazione interni al processo come le prove, le condizioni delle parti, eventuali precedenti ecc., ma anche a caratteri esterni del tutto riconducibili alla personalità del giudice ed avvenimenti riguardanti la vita dello stesso con coinvolgimento emotivo. Alcuni esempi, tutti afferenti al settore penale, pongono molti interrogativi. Nel testo pubblicato dal premio Nobel Daniel Kaheneman, insieme ai coautori Olivier Sibony, Cass R. Sunstein, dal titolo "Rumore. Un difetto del ragionamento umano"[9] si fa riferimento ad un esperimento[10] da cui emerge «che i giudici sono più propensi a concedere la libertà condizionale all'inizio della giornata o dopo la pausa pranzo che non immediatamente prima di una pausa. Un giudice affamato è più severo»[11] oppure «l'analisi condotta su sei milioni di sentenze formulate da giudici francesi nell'arco di dodici anni ha dimostrato che questi sono più clementi con gli imputati nel giorno del loro compleanno»[12]. Negli studi già citati si specifica che i principali vantaggi nell'affidarsi ad algoritmi, piuttosto che ad esseri umani, non risiede tanto

[8]C. CASTELLI, *giustizia predittiva,* in *questione giustizia,* rivista online*, 0*8/02/2022.

[9] La questione venne posta all'attenzione dell'opinione pubblica già a partire dai primi anni settanta dal giudice americano M. FRANKEL. A riguardo, si veda M. FRANKEL, *Criminal Sentences: Law Without Order*, Hil and Wang, New York 1973.

[10] Lo studio a cui Kahneman e gli altri autori del testo fanno riferimento è di S. DANZIGER, J. LEVAV & L. AVNAIM-PESSO, dal titolo *Extraneous Factors in Judicial Decisions,* "Proceedings of the National Academy of Sciences", 108, 2011, pp. 6889-6992 .

[11] D. KAHNEMAN, O. SIBONY, CASS.R SUNSTEIN, *Rumore. Un difetto del ragionamento umano,* 21 settembre 2021, UTET, Milano, op. cit. pag. 23.
Si veda anche A. GLÖCKNER, *The irrational hungry judge effect revisited: Simulations reveal magnitude of the effect is overestimated, "Judgment and Decision Making"*, vol. 11, no. 6, 2016, pp. 601-610. nello studio DANZIGER, LEVAV e AVNAIM-PESSO hanno analizzato le sentenze legali delle commissioni per la libertà vigilata israeliane, riguardanti l'effetto dell'ordine seriale attraverso cui i casi sono presentati all'interno delle sessioni di sentenza. Hanno scoperto che la probabilità di una decisione favorevole scende da circa il 65% a quasi lo 0% dalla prima decisione all'ultima decisione all'interno di ogni sessione e che il tasso di decisioni favorevoli ritorna al 65% in una sessione dopo una pausa alimentare. Gli autori sostengono che questi risultati forniscono supporto a fattori estranei che influenzano le decisioni giudiziarie e ipotizzano cautamente che l'effetto potrebbe essere guidato dall'esaurimento mentale.

[12] D. KAHNEMAN , O. SIBONY , CASS.R SUNSTEIN op. cit. pag.23.

nella superiorità delle regole quanto nella loro assenza di rumore[13] (c.d. "dispersione casuale"). In effetti, nel riprendere le parole emerse dal Report del Gruppo 1 Prevedibilità, predittività e umanità del giudicare della XIII Assemblea Nazionale degli Osservatori sulla giustizia civile: «Si è considerato che la "giustizia predittiva" incarna il mito illuminista del giudice bocca della legge, svelato ormai da gran tempo come tale nella manualistica della filosofia giuridica. E si sono evidenziate le anomalie di una giustizia siffatta: alla imparzialità del giudice, per darne attuazione in una declinazione mitologica, impossibile e distorta, si sostituisce l'incorporeità e la a-storicità di una macchina che ius dicit al di fuori della storia, cioè lo spazio abitato dagli umani e cioè dai loro corpi. Ci troviamo forse di fronte al recupero in chiave tecnocratica di una teocratica "Giustizia Eterna"? O forse solo a un alibi de-responsabilizzante? Comunque, un giudizio che è specchio della profonda alienazione di un Essere Umano che delega il giudizio su un altro uomo a qualcosa di non umano. Rimane, per ora, la clausola di garanzia finale recata dalla norma sopra indicata, ma limitata: è garantito l'intervento umano, non chiare le sue modalità e la certa sua decisività. Ma rimane anche il problema, ampiamente evidenziato, del rapporto tra la decisione del giudice e quella della macchina, tutto da verificare e scoprire, anche a livello di legittimazione diffusa nelle decisioni»[14]. Ciò che realmente emerge dalle considerazioni anzidette è che non esiste una linea univoca sull'impiego di nuove tecnologie di giustizia predittiva, ma è indispensabile un approccio evolutivo e progressivo a quella che può essere considerata una vera e propria rivoluzione copernicana nella gestione del comparto giustizia, soprattutto in ambito penale. La chiave di volta, probabilmente, risiede in un

---

[13] D. KAHNEMAN , O. SIBONY , CASS.R SUNSTEIN op. cit. pag.15.

[14] R*eport* su *"Prevedibilità, predittività e umanità del giudicare"* della XIII *Assemblea Nazionale degli Osservatori sulla Giustizia Civile "Equilibrio tra processo e autonomia privata"*, Reggio Emilia 8, 9 e 10 giugno 2018.
Ulteriori perplessità nascono da esperienze di altri paesi. Il sito *Justice predictive* nato in ambito forense in Francia dava agli avvocati la possibilità, con accesso a pagamento, di sapere gli orientamenti dei diversi Tribunali con una stima delle probabilità di successo di una causa avanzata davanti a quel Foro. Questa ottica probabilistica delle decisioni potrebbe portare a abnormità quali verificare le percentuali degli esiti di un procedimento in cui vi sia il binomio tra un certo magistrato ed un certo avvocato, classificare il magistrato, puntare sull'avvocato che con quel magistrato ha vinto più cause. Tutti dati che potrebbero alterare e condizionare sia la scelta del difensore che le stesse decisioni e incidevano sull'immagine dei diversi attori del processo, sulla professionalità degli avvocati e sull'indipendenza del giudicante. Tant'è che in Francia, dove le sperimentazioni sono più avanzate, la profilazione di magistrati e avvocati è stata vietata con una norma munita di sanzione penale. Si veda C. CASTELLI, *giustizia predittiva,* in Q*uestione Giustizia,* rivista *online*, 08/02/2022.

atteggiamento pacato ed esplorativo di un fenomeno che, per quanto avversato dalla dottrina e da parti consistenti degli addetti ai lavori, offre alcune opportunità meritevoli di considerazione, anche alla luce dei potenziali rischi che una simile tecnologia potrebbe comportare nel caso in cui sia lasciata a sé stessa o, peggio ancora, a mani poco esperte. E' fondamentale comprenderne i meccanismi e, come per ogni altra precedente evoluzione della tecnica, non è facile eliminare completamente gli effetti di un processo in divenire, semmai, si può tentare di mitigarne le criticità. Tutto ciò richiede attenzione, interesse e predisposizione ad approcciare con spirito critico e non aprioristicamente contrario.

3. Esempi concreti d' impiego (il caso Loomies).

Il miglior modo per garantire una adeguata consapevolezza del potenziale delle moderne tecnologie di A.I. è studiarne i principali esperimenti di impiego, già concretizzatisi in alcuni sistemi giudiziari d' oltreoceano. Inutile far riferimento ad ambiti poco agevoli e carichi di opacità tipici dei regimi totalitari. In tali contesti, come quello cinese, seppur non mancano esempi di sperimentazione concreta di A.I. in ambito penale, almeno per il momento, è assai difficile, se non impossibile, recuperare informazioni meritevoli di considerazione alla luce della scarsa genuinità o completezza intrinseca dei dati[15]. Più agevole, anche per le ripercussioni già avute nei circoli di studio della materia o, a maggior ragione, nella costruzione di proposte normative in ambito europeo, è il caso di impiego del cosiddetto sistema COMPAS. Acronimo di Correctional offender management profiling for alternative

[15] L'esempio tipico è quello del procuratore *robot* nell'ambito del processo penale per la valutazione delle prove, dei presupposti per custodia cautelare e arresto e della pericolosità dell'indagato/imputato. Il sistema, sviluppato da un *equipe* di ricercatori sotto la guida del Prof. Shi Young utilizza per la valutazione un *database* di 17000 casi avvenuti tra il 2015 ed il 2020, ma, al momento, è in grado di analizzare solo otto fattispecie di reato previste dal Codice penale cinese (nella fattispecie frodi con carte di credito, gioco d'azzardo, guida pericolosa, lesioni internazionali, intralcio ai doveri d'ufficio, furto, frode e "*scelta di litigi e provocazione di guai*"). Secondo quanto dichiarato dal Prof. Young, dai primi *test* emergerebbe un margine di errore del 3% nella valutazione del grado colpevolezza od innocenza dell'imputato, ma senza che il *robot* possa prendere parte alla parte attiva della decisione. Lo stesso interverrebbe come mero ausilio al libero convincimento del giudice. Infatti, sempre secondo il Prof. Young, affinchè la macchina possa prendere autonomamente decisioni avrebbe bisogno di "*convertire un linguaggio umano complesso ed in continua evoluzione in un formato matematico o geometrico standard che un computer potrebbe capire*". Per ulteriori approfondimenti si veda l'articolo pubblicato sul *South China morning Post* il 26 dicembre 2021 al seguente link: https://www.scmp.com/news/china/science/article/3160997/*chinese-scientists-develop-ai-prosecutor-can-press-its-own*.

sanctions, è un algoritmo coperto da brevetto e sviluppato da un'azienda privata di nome Equivant, presentato come strumento di valutazione della possibilità di recidiva di un imputato sulla base di una serie di risposte ad un questionario standard di 137 quesiti. Si tratta di uno dei tanti sistemi di valutazione del rischio a disposizione dell'apparato giudiziario statunitense, impiegato per ricevere informazioni utili a valutare la possibilità di reati futuri e le tipologie di misure da applicare nella prevenzione[16]. Il caso più interessante nell'impiego dell'algoritmo è Supreme Court of Wisconsin, State of Wisconsin v. Eric L. Loomis, Case no. 2015AP157-CR, 5 April – 13 July 2016, in cui la Corte Suprema del Wisconsin (State o Wisconsin v. Eric L. Loomis, 13 Luglio 2016) si è pronunciata sull'appello del sig. Eric L. Loomis, la cui pena a sei anni di reclusione era stata comminata dal Tribunale circondariale di La Crosse. Nel determinare la pena, i giudici, avevano tenuto conto dei risultati elaborati dal programma COMPAS secondo cui Loomis era da considerarsi soggetto ad alto rischio di recidiva. Riassumendo brevemente il fatto, nel 2013, Loomies è stato fermato alla guida di un'auto precedentemente impiegata in una sparatoria. All'esito del fermo, gli vennero contestati cinque capi di accusa in recidiva: messa in pericolo della sicurezza, tentativo di fuga o elusione di un ufficiale del traffico, guida di un veicolo senza consenso del proprietario, possesso di arma da fuoco da parte di un pregiudicato e possesso di fucile a canna corta o pistola. All'esito di patteggiamento, l'imputato ha concordato la pena per le accuse meno severe di tentativo di fuga ed elusione di un ufficiale del traffico. Allo stesso tempo, la corte, dopo l'accoglimento di ammissione di colpevolezza di Loomis, ha ordinato una cosiddetta Presentence Investigation Report (PSI). Per il sistema statunitense,

[16] «Per "*polizia predittiva*" possiamo intendere l'insieme delle attività rivolte allo studio e all'applicazione di metodi statistici con l'obiettivo di "*predire*" chi potrà commettere un reato, o dove e quando potrà essere commesso un reato, al fine di prevenire la commissione dei reati stessi. La predizione si basa fondamentalmente su una rielaborazione attuariale di diversi tipi di dati, tra cui quelli relativi a notizie di reati precedentemente commessi, agli spostamenti e alle attività di soggetti sospettati, ai luoghi, teatro di ricorrenti azioni criminali, e alle caratteristiche di questi luoghi, al periodo dell'anno o alle condizioni atmosferiche maggiormente connesse alla commissione di determinati reati; tra i dati utilizzati a questi fini talora compaiono anche informazioni relative all'origine etnica, al livello di scolarizzazione, alle condizioni economiche, alle caratteristiche somatiche (...una rivincita di Lombroso?), riconducibili a soggetti appartenenti a determinate categorie criminologiche (ad es., potenziali terroristi), etc». Così F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *diritto penale e uomo, rivista,* 29 settembre 2019. Per un completo inquadramento della materia della predictive policing, v. W.L. PERRY, B. MCINNIS, C.C. PRICE, S.C. SMITH, J.S. HOLLYWOOD, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, Rand Corporation*, 2013.

si tratta in pratica di una relazione stesa sulla base dei risultati ottenuti dalle attività investigative in merito alla storia personale dell'imputato, il cui svolgimento è inserito in una fase preliminare alla sentenza sulla determinazione della pena; l'equivalente del nostro giudizio di conformità del giudice, il cui fine ultimo era quello di modularne gli esiti in base ai risultati raggiunti. Nel caso specifico la PSI ha integrato i risultati del sistema COMPAS. Il sistema non è altro che uno strumento concepito sia per prevedere il potenziale rischio di recidiva, sia per identificare i bisogni attuali dell'individuo in aree specifiche quali occupazione, disponibilità di alloggio ed abuso di sostanze stupefacenti. L'algoritmo di funzionamento del sistema non fa altro che elaborare i dati ottenuti dal fascicolo dell'imputato e dalle risposte ricevute dallo stesso in sede di colloqui, ottenendo così un grafico in tre barre in cui si evince, in una scala da 1 a 10, il rischio di recidiva preprocessuale, di recidiva generale e di recidiva violenta. Tali punteggi hanno la funzione di quantificare in termini numerici la probabilità che i soggetti con storia di precedenti simili siano più o meno propensi a compiere nuovamente reati una volta tornati in libertà[17]. Nel caso specifico di Loomis, il sistema di calcolo indicava una valutazione di alto livello di rischio in tutti e tre gli ambiti, anche se il PSI indicava che il risultato dell'algoritmo non dovesse essere impiegato per decidere l'intensità della pena o la possibilità di reclusione. Nonostante ciò, il tribunale circondariale, nel determinare la pena, ponderando tra i vari fattori anche i risultati di COMPAS, ha deciso di negare la libertà vigilata, tenendo in considerazione, tra gli altri, fattori quali gravità del crimine, condotta dell'imputato in libertà ed in custodia, nonché ulteriori capi di accusa per i quali lo stesso Loomis non aveva la cosiddetta "read in charges", ossia, una dichiarazione di colpevolezza. All'esito della decisione del tribunale, Loomis ha deciso di presentare istanza per lesione ai propri diritti ad un equo processo dall'uso delle determinazioni del sistema COMPAS[18]. Il tribunale circondariale ha rigettato, comunque, l'istanza,

[17] L'aspetto fondamentale da tenere in considerazione è che il sistema *COMPAS* non prevede il rischio di recidiva individuale dell'imputato, bensì elabora tale previsione sulla base di una comparazione delle informazioni ottenute dal singolo *PSI* con quelle relative ad un gruppo di individui con caratteristiche assimilabili.

[18] «Nella seconda udienza post-condanna, la difesa di Loomis chiamava a testimoniare un esperto, il quale dichiarava che lo strumento *COMPAS* non era stato concepito per l'utilizzo nelle decisioni di incarcerazione. Secondo il consulente, la corte che prende in considerazione i risultati forniti dall'algoritmo incorre fortemente nella probabilità di sovrastimare il rischio di recidiva individuale e di determinare la pena dell'imputato sulla scorta di fattori ininfluenti. Inoltre, affermava che i tribunali posseggono poche informazioni sul processo di analisi del rischio effettuato da *COMPAS*, non sapendo ad esempio come il sistema compia la comparazione della storia individuale dell'imputato con quella del gruppo di popolazione preso a riferimento, né se tale gruppo di individui appartenga al medesimo stato americano».

adducendo l'inevitabiltà della decisione medesima, anche in assenza delle specifiche determinazioni dell'algoritmo. Dopo rimessione del tribunale inferiore e, giunto alla Corte Suprema del Wisconsin, l'impianto difensivo ha costruito un paradigma basato su tre principali assunti, tutti strettamente connessi alla violazione del diritto ad un equo processo. Nello specifico, sono stati sollevati dubbi circa 1) il diritto ad essere condannato ad una pena determinata in virtù di informazioni accurate e precise, delle quali, però, nel caso di impiego dell'algoritmo in questione, non si poteva disporre, in quanto coperte da diritti di proprietà industriale della società; 2) il diritto di essere condannato ad una pena individualizzata; 3) l'uso improprio del dato costruito sull'appartenenza ad un genere specifico (il fatto di essere nero)[19], nella determinazione della pena. Nel suo giudizio, la Corte ha ribadito come l'algoritmo possa essere in linea teorica impiegato nei giudizi di determinazione della pena, ma, allo stesso tempo, richiede un uso particolarmente limitato ed oggetto di cautela. Come corollario ulteriore, i giudici hanno stabilito che i dati ricavati dal sistema COMPAS non possono

---

di veda S. CARRER, *Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin,* su G*iurisprudenza Penale*, sito *online*, 24 aprile 2019.

[19] E' interessante notare come il *GDPR* europeo, al considerando 51, affermi come «meritano una specifica protezione i dati personali che, per loro natura, sono particolarmente sensibili sotto il profilo dei diritti e delle libertà fondamentali, dal momento che il contesto del loro trattamento potrebbe creare rischi significativi.... Tra tali dati personali dovrebbero essere compresi anche i dati personali che rivelano l'origine razziale o etnica, essendo inteso che l'utilizzo dei termini «origine razziale», nel presente regolamento, non implica l'accettazione da parte dell'Unione di teorie che tentano di dimostrare l'esistenza di razze umane distinte. Il trattamento di fotografie non dovrebbe costituire sistematicamente un trattamento di categorie particolari di dati personali, poiché esse rientrano nella definizione di dati biometrici soltanto quando siano trattate attraverso un dispositivo tecnico specifico che consente l'identificazione univoca o l'autenticazione di una persona fisica. Tali dati personali non dovrebbero essere oggetto di trattamento, a meno che il trattamento non sia consentito nei casi specifici di cui al presente regolamento, tenendo conto del fatto che il diritto degli Stati membri può stabilire disposizioni specifiche sulla protezione dei dati per adeguare l'applicazione delle norme del presente regolamento ai fini della conformità a un obbligo legale o dell'esecuzione di un compito di interesse pubblico o per l'esercizio di pubblici poteri di cui è investito il titolare del trattamento. Oltre ai requisiti specifici per tale trattamento, dovrebbero applicarsi i principi generali e altre norme del presente regolamento, in particolare per quanto riguarda le condizioni per il trattamento lecito. È opportuno prevedere espressamente deroghe al divieto generale di trattare tali categorie particolari di dati personali, tra l'altro se l'interessato esprime un consenso esplicito o in relazione a esigenze specifiche, in particolare se il trattamento è eseguito nel corso di legittime attività di talune associazioni o fondazioni il cui scopo sia permettere l'esercizio delle libertà fondamentali».

comunque essere determinanti nella decisione finale. L'impossibilità di confutazione scientifica dell'algoritmo in quanto coperto da brevetto e segreto industriale non ha però inibito la Corte nell'affermare il principio per cui, "anche in assenza di accesso al meccanismo di funzionamento del sistema, l'imputato avesse la materiale possibilità di contestare i risultati finali di calcolo del rischio"[20]. Riguardo al terzo punto, nonostante la produzione di alcuni studi da parte della difesa concernenti il livello di affidabilità dell'algoritmo, la Corte ha, tuttavia, stabilito che le esperienze nell'impiego del sistema, già maturate in altri Stati, fossero sufficienti a supportare un notevole grado di affidabilità, ribadendo, però, la necessaria garanzia di accuratezza di COMPAS, mediante un costante monitoraggio e aggiornamento del sistema sulla base delle evoluzioni riguardanti i dati ritenuti più critici, legati essenzialmente a fattori sociali, familiari e di appartenenza etnica, nonché la presenza di pregresse esperienze criminali. La possibile generale tendenza del software ad attribuire rischi di recidiva a particolari gruppi sociali, senza prendere in considerazione le caratteristiche individuali dell'imputato, è stata ridimensionata dal giudizio della Corte col presupposto giuridico di necessario uso della discrezionalità da parte dell'organo giudicante, il quale si troverebbe a dover tenere conto non di un singolo fattore, ma della loro integrità complessiva. L'eccezione di incostituzionalità, sollevata da Loomis, basata sulla presunta discriminazione di genere compiuta da COMPAS nell'attribuire un grado di maggior rischio di recidiva negli individui maschi, non è stata ritenuta dalla Corte sufficiente poiché il giudizio di primo grado aveva tenuto conto di numerosi altri fattori e, in virtù dell'esame complessivo della vicenda, infine, è stata ribadita l'assenza di violazione al diritto ad un equo processo. Il caso ha consentito la definizione dei parametri di legittimità nell'impiego del software stabilendo che sistemi come COMPAS possano essere ritenuti rilevanti solo in questioni inerenti alla comminazione di misure alternative alla detenzione per gli individui a basso rischio di recidiva, nella valutazione della possibilità di sottoporre un criminale a controllo in modo sicuro all'interno della società, anche con misure alternative alla detenzione come l'affidamento in prova, nonché nell'imposizione di termini e condizioni per la concessione della libertà vigilata o per le eventuali sanzioni alle violazioni di regole previste dai regimi alternativi alla detenzione. Di fatto, ribadendo il necessario esercizio discrezionale del giudice nella valutazione dei vari fattori in gioco, la Corte

[20] Nonostante la segretezza del processo di funzionamento del *software*, dal manuale del sistema *COMPAS* si evince come il meccanismo di assegnazione dei punteggi di rischio sia basato essenzialmente su processi statistici (relativi a condotta criminale), con uso limitato di variabili dinamiche (come ad esempio contesto sociale, uso di droghe). Il *report* di *COMPAS* contiene una lista di 21 domande e risposte riguardanti fattori statistici.

ha specificato l'impossibilità di un impiego del sistema nella determinazione della pena o sulla scarcerazione dell'imputato. In concreto, l'unica finalità dell'algoritmo in questione è quella di valutare il rischio di recidiva[21].

4. L' approccio europeo al tema (trasparenza e proporzione).

Come ogni altro fenomeno influenzato dall'evoluzione tecnologica, è fondamentale porre un notevole grado di attenzione sia sui potenziali rischi che sulle innegabili opportunità. Quando si discute di giustizia predittiva, se da una parte è innegabile il raggiungimento di obiettivi pratici, rappresentati principalmente dal risparmio in termini di costo e tempo necessari a processare operazioni di ricostruzione dei fatti estremamente complesse, dall'altro è innegabile il potenziale rischio di una significativa vulnerabilità nel sistema di garanzie e tutele dell'individuo, nonché sull'esercizio della funzione giurisdizionale. Ci si è chiesti se casi analoghi al tipo COMPAS possano un giorno approdare anche in ambito europeo. Di fatto, l'UE ha già avviato un percorso di risposta ai rischi connessi al processo di evoluzione tecnologica con lo scudo del GDPR, uno strumento regolamentare costruito su modelli di prevenzione e flessibilità in virtù di probabili cambiamenti futuri[22]. Come specificato dal regolamento nel punto 6 dei considerando, «la rapidità dell'evoluzione tecnologica e la globalizzazione comportano nuove sfide per la protezione dei dati personali. La portata della condivisione e della raccolta di dati personali è aumentata in modo significativo. La tecnologia attuale consente tanto alle imprese private quanto alle autorità pubbliche di utilizzare dati personali, come mai in precedenza, nello svolgimento delle loro attività. Sempre più spesso, le persone fisiche rendono disponibili al pubblico su scala mondiale informazioni personali che le riguardano. La

[21] Secondo "*l'opinione concorrente*" del giudice Drake Roggensack, «i giudici sono legittimati a considerare ("*consider*") i dati forniti dal *software* nella determinazione della sentenza, insieme però ad una moltitudine di altri fattori. Illegittimo sarebbe invece basare ("*rely*") la sentenza su tali risultati, utilizzandoli quindi come fattori determinanti della decisione». Al riguardo S. CARRER, *Se l'*amicus curiae *è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in *Giurisprudenza Penale Web*, 2019.

[22] Al considerando 129 *GDPR* si dice che «Al fine di garantire un monitoraggio e un'applicazione coerenti del presente regolamento in tutta l'Unione, le autorità di controllo dovrebbero avere in ciascuno Stato membro gli stessi compiti e poteri effettivi, fra cui poteri di indagine, poteri correttivi e sanzionatori, e poteri autorizzativi e consultivi, segnatamente in caso di reclamo proposto da persone fisiche, e fatti salvi i poteri delle autorità preposte all'esercizio dell'azione penale ai sensi del diritto degli Stati membri, il potere di intentare un'azione e di agire in sede giudiziale o stragiudiziale in caso di violazione del presente regolamento. Tali poteri dovrebbero includere anche il potere di imporre una limitazione provvisoria o definitiva al trattamento, incluso il divieto di trattamento.».

tecnologia ha trasformato l'economia e le relazioni sociali e dovrebbe facilitare ancora di più la libera circolazione dei dati personali all'interno dell'Unione e il loro trasferimento verso paesi terzi e organizzazioni internazionali, garantendo al tempo stesso un elevato livello di protezione dei dati personali». E ancora, al punto 4, «Il trattamento dei dati personali dovrebbe essere al servizio dell'uomo. Il diritto alla protezione dei dati di carattere personale non è una prerogativa assoluta, ma va considerato alla luce della sua funzione sociale e va contemperato con altri diritti fondamentali, in ossequio al principio di proporzionalità». Nonostante le sue lodevoli intenzioni, lo stesso GDPR non è in grado di stare al passo dei progressi in ambito tecnologico, richiedendo un costante sforzo interpretativo e di adeguamento. Con riferimento al caso oggetto di approfondimento, il GDPR riserva nell' art. 22 uno specifico dettato normativo relativo al divieto di essere sottoposti a decisioni interamente automatizzate: così al comma 1 si specifica che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona»[23]. Ciò significa che, qualora dovesse intervenire un algoritmo nella

[23] Così, ai considerando collegati 71 e 72 «l'interessato dovrebbe avere il diritto di non essere sottoposto a una decisione, che possa includere una misura, che valuti aspetti personali che lo riguardano, che sia basata unicamente su un trattamento automatizzato e che produca effetti giuridici che lo riguardano o incida in modo analogo significativamente sulla sua persona. Tale trattamento comprende la «profilazione», che consiste in una forma di trattamento automatizzato dei dati personali che valuta aspetti personali concernenti una persona fisica, in particolare al fine di analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze o gli interessi personali, l'affidabilità o il comportamento, l'ubicazione o gli spostamenti dell'interessato, ove ciò produca effetti giuridici che la riguardano o incida in modo analogo significativamente sulla sua persona. In ogni caso, tale trattamento dovrebbe essere subordinato a garanzie adeguate, che dovrebbero comprendere la specifica informazione all'interessato e il diritto di ottenere l'intervento umano, di esprimere la propria opinione, di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la decisione. Tale misura non dovrebbe riguardare un minore. Al fine di garantire un trattamento corretto e trasparente nel rispetto dell'interessato, tenendo in considerazione le circostanze e il contesto specifici in cui i dati personali sono trattati, è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, metta in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e che impedisca tra l'altro effetti discriminatori nei confronti di persone fisiche sulla base della razza

decisione, ma la stessa sia sottoposta al controllo umano, lo scudo previsto dall'articolo in questione non sarebbe applicabile, perdendo di valore tutelante. L'imputato avrebbe sempre diritto a ricevere informazione sull'esistenza di un processo decisionale automatizzato e sui meccanismi alla base di tale funzionamento, incluse le potenziali conseguenze giuridiche del suo impiego. Al riguardo gli art. 13-14-15 del GDPR potrebbero rappresentare una forma di tutela in caso di decisioni in cui sia stato coinvolto l'uso di un algoritmo. Al paragrafo 1 dell'art 13 è disposto che «In caso di raccolta presso l'interessato di dati che lo riguardano, il titolare del trattamento fornisce all'interessato, nel momento in cui i dati personali sono ottenuti, informazioni riguardanti: a) l'identità e i dati di contatto del titolare del trattamento e, ove applicabile, del suo rappresentante; b) i dati di contatto del responsabile della protezione dei dati, ove applicabile; c) le finalità del trattamento cui sono destinati i dati personali nonché la base giuridica del trattamento; d) qualora il trattamento si basi sull'articolo 6, paragrafo 1, lettera f), i legittimi interessi perseguiti dal titolare del trattamento o da terzi; e) gli eventuali destinatari o le eventuali categorie di destinatari dei dati personali; f) ove applicabile, l'intenzione del titolare del trattamento di trasferire dati personali a un paese terzo o a un'organizzazione internazionale...». E ancora, al paragrafo 2, che il titolare del trattamento fornisce all'interessato le ulteriori informazioni necessarie per garantire un trattamento corretto e trasparente nonché, alla lettera f), «l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, l'importanza e le conseguenze previste da tale trattamento per l'interessato»[24] . Anche l'art. 15, al paragrafo 1, lettera f) dispone «il diritto di proporre reclamo a un'autorità di controllo; g) qualora i dati non siano raccolti presso

---

o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero che comportano misure aventi tali effetti. Il processo decisionale automatizzato e la profilazione basati su categorie particolari di dati personali dovrebbero essere consentiti solo a determinate condizioni».

[24] I considerando 60 e 61 indicano che «I principi di trattamento corretto e trasparente implicano che l'interessato sia informato dell'esistenza del trattamento e delle sue finalità. Il titolare del trattamento dovrebbe fornire all'interessato eventuali ulteriori informazioni necessarie ad assicurare un trattamento corretto e trasparente, prendendo in considerazione le circostanze e del contesto specifici in cui i dati personali sono trattati. Inoltre, l'interessato dovrebbe essere informato dell'esistenza di una profilazione e delle conseguenze della stessa. In caso di dati personali raccolti direttamente presso l'interessato, questi dovrebbe inoltre essere informato dell'eventuale obbligo di fornire i dati personali e delle conseguenze in cui incorre se si rifiuta di fornirli».

l'interessato, tutte le informazioni disponibili sulla loro origine; h) di comunicare l'esistenza di un processo decisionale automatizzato, compresa la profilazione...». La disciplina relativa all'esistenza di un processo decisionale automatizzato è ulteriormente approfondita al considerando 63 in cui è specificato che «un interessato dovrebbe avere il diritto di accedere ai dati personali raccolti che lo riguardano e di esercitare tale diritto facilmente e a intervalli ragionevoli, per essere consapevole del trattamento e verificarne la liceità». Ogni interessato, inoltre, «dovrebbe pertanto avere il diritto di conoscere e ottenere comunicazioni in particolare in relazione alla finalità per cui i dati personali sono trattati, ove possibile al periodo in cui i dati personali sono trattati, ai destinatari dei dati personali, alla logica cui risponde qualsiasi trattamento automatizzato dei dati e, almeno quando è basato sulla profilazione, alle possibili conseguenze di tale trattamento. Ove possibile, il titolare del trattamento dovrebbe poter fornire l'accesso remoto a un sistema sicuro che consenta all'interessato di consultare direttamente i propri dati personali. Tale diritto non dovrebbe ledere i diritti e le libertà altrui...». In base al costrutto normativo previsto dagli articoli in questione, l'interessato avrebbe diritto a ricevere tutta una serie di informazioni relative all'esistenza di un processo decisionale automatizzato[25], sulla tecnica alla base del suo funzionamento e, soprattutto, sulle conseguenze derivanti dal suo impiego. I principi relativi alle coperture del segreto industriale, indicano, però, che lo stesso non può avere accesso totale alle formule matematiche impiegate dall'algoritmo, ma deve comunque avere accesso a tutta una serie di informazione che consentano una comprensione semplice ed intelligibile del suo funzionamento[26]. Allo stesso tempo, però, il GDPR, prevede una serie di deroghe specifiche, coperte da riserva di legge, in cui può essere limitata la portata degli obblighi derivanti dal regolamento, a condizione di rispettare i parametri essenziali delle libertà fondamentali e che sia una misura necessaria e proporzionata in una società democratica. Più nello specifico, l'art. 23 prevede come elementi giustificanti di deroga la presenza di esigenze riguardanti: «a) la sicurezza nazionale; b) la difesa; c) la sicurezza pubblica; d)

[25] I primi accenni della giurisprudenza italiana in tal senso provengono dal settore del diritto amministrativo. Una recente Sentenza del Consiglio di Stato, la n. 2270 dell'8 aprile 2019, sez. VI, ha legittimato l'impiego di algoritmi nelle procedure valutative dalla pubblica amministrazione, a condizione, però, che vi sia trasparenza e possibilità di controllo giurisdizionale. I giudici hanno precisato che l'impiego di algoritmi e procedure automatizzate è considerato un "a*tto amministrativo informatico*" e, come tale, deve necessariamente sottostare ai principi generali di ragionevolezza, pubblicità, proporzionalità e trasparenza.

[26] Il Parlamento europeo, nel *report* del 2019 *AI and Robotics,* ha specificato la necessità di avere la c.d. "*intellligibility of decisions",* nonché il diritto della persona interessata dal trattamento dei dati di essere informato sulla logica del meccanismo di elaborazione e sulla presenza di un controllo umano esterno.

la prevenzione, l'indagine, l'accertamento e il perseguimento di reati o l'esecuzione di sanzioni penali, incluse la salvaguardia contro e la prevenzione di minacce alla sicurezza pubblica»[27]. Il fenomeno, in continua evoluzione, non ha lasciato indifferente l'unione europea, la quale, in un quadro normativo non agevole, attraverso la commissione, ha iniziato un percorso di regolamentazione con una proposta di armonizzazione delle regole in ambito europeo. Come si può leggere sul sito ufficiale, «the European approach to artificial intelligence (A.I.) will help build a resilient Europe for the Digital Decade where people and businesses can enjoy the benefits of A.I. It focuses on 2 areas: excellence in A.I. and trustworthy A.I. The European approach to A.I. will ensure that any A.I. improvements are based on rules that safeguard the functioning of markets and the public sector, and people's safety and fundamental rights»[28]. Si può facilmente comprendere come l'iniziativa europea sia essenzialmente orientata ad un contenimento degli aspetti critici legati all'uso della tecnologia di A.I., soprattutto con riguardo alla tutela delle libertà fondamentali, senza, però, trascurarne i potenziali elementi di carattere positivo[29]. In generale, gli Stati membri prevedono un incremento delle risorse da investire in tale settore. Secondo l'approccio strategico alla A.I., «maximising resources and coordinating investments is a critical component of the Commission's A.I. strategy. Through the Digital Europe and Horizon Europe programmes, the

[27] Infatti, nel considerando 73 è chiarito che «Il diritto dell'Unione o degli Stati membri può imporre limitazioni a specifici principi e ai diritti di informazione, accesso, rettifica e cancellazione di dati, al diritto alla portabilità dei dati, al diritto di opporsi, alle decisioni basate sulla profilazione, nonché alla comunicazione di una violazione di dati personali all'interessato e ad alcuni obblighi connessi in capo ai titolari del trattamento, ove ciò sia necessario e proporzionato in una società democratica per la salvaguardia della sicurezza pubblica, ivi comprese la tutela della vita umana, in particolare in risposta a catastrofi di origine naturale o umana, le attività di prevenzione, indagine e perseguimento di reati o l'esecuzione di sanzioni penali, incluse la salvaguardia contro e la prevenzione di minacce alla sicurezza pubblica...».

[28]Si veda il sito ufficiale dell'Unione Europea alla pagina https://*digital strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence*.

[29] Nell' a*nnual report lays out the challenges of protecting fundamental rights in the digital age* del 10 Dicembre 2021 è specificato che *«the increasing use of artificial intelligence systems can yield great benefits, but certain applications are complex and opaque, which can be a challenge for compliance with or enforcement of fundamental rights. Many Member States have developed national strategies on artificial intelligence to ensure transparency, traceability and robustness and find effective ways to comply with fundamental rights. In April 2021, the Commission proposed a legislative act to ensure that artificial intelligence systems that pose a high-risk to fundamental rights are appropriately tested and documented»*. Per approfondimenti, si veda il *report* sul sito ufficiale della commissione europea.

Commission plans to invest €1 billion per year in A.I. It will mobilise additional investments from the private sector and the Member States in order to reach an annual investment volume of €20 billion over the course of the digital decade[30]». Il raggiungimento di una elevata qualità del sistema è ritenuto un fattore determinante nella costruzione del sistema di gestione. Non solo, «Initiatives such as the EU Cybersecurity Strategy, the Digital Services Act and the Digital Markets Act, and the Data Governance Act provide the right infrastructure for building such systems». Affinché si possa avere fiducia nei sistemi di intelligenza artificiale, la commissione europea ha proposto tre iniziative legislative convergenti in tal senso. Si parla di a) European legal framework for A.I. to address fundamental rights and safety risks specific to the A.I. systems; b) EU rules to address liability issues related to new technologies, including A.I. systems ; c) a revision of sectoral safety legislation (e.g. Machinery Regulation, General Product Safety Directive). Secondo la commissione, una mitigazione dei rischi connessi all'uso dell'A.I. può avvenire mediante una legislazione complementare con norme proporzionate e flessibili. In tale modo, l'unione europea dovrebbe divenire un riferimento gold standard globale nel settore dell'intelligenza artificiale. Sempre secondo la commissione, con un simile quadro di riferimento, sviluppatori ed utenti della tecnologia avrebbero un quadro giuridico preciso basato su quattro diversi livelli di rischio: rischio "inaccettabile**"**, rischio "elevato**"**, rischio "limitato**"** e rischio "minimo**"**. Nella relazione alla proposta di regolamento si può leggere una prima definizione di intelligenza artificiale, intendendo con la stessa "una famiglia di tecnologie in rapida evoluzione in grado di apportare una vasta gamma di benefici economici e sociali in tutto lo spettro delle attività industriali e sociali". Tenendo in considerazione la velocità del progresso tecnologico, con la proposta di regolamento l'UE si è impegnata a perseguire un approccio equilibrato cercando di preservare «la leadership tecnologica dell'UE e assicurare che i cittadini europei possano beneficiare di nuove tecnologie sviluppate e operanti in conformità ai valori, ai diritti fondamentali e ai principi dell'Unione»[31]. La costruzione di un modello di

[30]Si veda il sito ufficiale dell'unione europea alla pagina https*://digital strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence*

[31] Come si legge la paragrafo 1.1 della relazione alla proposta con la stessa «si tiene fede all'impegno politico della presidente Von Der Leyen che, nei suoi orientamenti politici per la Commissione 2019-2024 "*Un'Unione più ambiziosa*", ha annunciato che la Commissione avrebbe presentato una normativa per un approccio europeo coordinato alle implicazioni umane ed etiche dell'intelligenza artificiale. A seguito di tale annuncio la Commissione ha pubblicato il 19 febbraio 2020 il *Libro bianco sull'intelligenza artificiale* - Un approccio europeo all'eccellenza e alla fiducia. Il Libro bianco definisce le opzioni strategiche su come conseguire il duplice obiettivo di promuovere l'adozione dell'*A.I.* e affrontare i rischi associati a determinati utilizzi di tale tecnologia».

gestione umanocentrica della tecnologia di intelligenza artificiale, inoltre, la si evince dalla specificazione secondo cui «Le regole per l'A.I. disponibili sul mercato dell'Unione o che comunque interessano le persone nell'Unione dovrebbero pertanto essere incentrate sulle persone, affinché queste ultime possano confidare nel fatto che la tecnologia sia usata in modo sicuro e conforme alla legge, anche in termini di rispetto dei diritti fondamentali». In alcune recenti conclusioni del consiglio europeo riguardo all'uso dell'intelligenza artificiale si è indicato, come requisito per il rispetto dei diritti fondamentali e l'agevolazione nell'applicazione delle norme di riferimento, di tenere sempre in considerazione l'opacità, la complessità, la faziosità ed un certo grado di imprevedibilità nel comportamento di taluni sistemi di A.I.[32]. Anche il parlamento europeo è stato caratterizzato da un incremento della attività nel settore A.I. mediante tutta una serie di risoluzioni, anche in ambito penale[33]. Relativamente ai profili più critici quali protezione dei dati, non discriminazione, nella proposta della commissione non mancano elementi di attenzione specifici. Più nel dettaglio, al paragrafo 1.2 della relazione è indicato come la proposta assicuri «la coerenza con la Carta dei diritti fondamentali dell'Unione europea e il diritto derivato dell'UE in vigore in materia di protezione dei dati, tutela dei consumatori, non discriminazione e parità di genere». La proposta tende anche a non pregiudicare «la direttiva sulla protezione dei dati nelle attività di polizia e giudiziarie (direttiva (UE) 2016/680) e li integra con una serie di regole armonizzate applicabili alla progettazione, allo sviluppo e all'utilizzo di determinati sistemi di A.I. ad alto rischio nonché di restrizioni concernenti determinati usi dei sistemi di identificazione biometrica remota». Inoltre, sono previsti tutta una serie di criteri volti a ridurre al minimo il rischio di "discriminazione algoritmica", con riferimento specifico al momento della progettazione e, in particolar modo, alla qualità dei dati impiegati nel settaggio dell'algoritmo, «integrati con obblighi relativi alle prove, alla gestione dei rischi, alla documentazione e alla sorveglianza umana durante l'intero ciclo di vita dei sistemi di A.I.»[34].

[32] Consiglio dell'Unione Europea, *Conclusioni della presidenza – La Carta dei diritti fondamentali nel contesto dell'intelligenza artificiale e della trasformazione digitale*, 11481/20, 2020.

[33] A riguardo si veda Progetto di relazione del Parlamento europeo sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale, 2020/2016 (INI).

[34] Al termine del paragrafo 1.2 della relazione alla proposta è specificato che «In relazione ai sistemi di *A.I.* che sono componenti di sistemi informatici su larga scala nello spazio di libertà, sicurezza e giustizia gestiti dall'Agenzia dell'Unione europea per la gestione operativa dei sistemi *IT* su larga scala (*eu-LISA*), la proposta non si applicherà ai sistemi di *A.I.* immessi sul mercato o messi in servizio prima che sia trascorso un anno dalla data di applicazione del presente regolamento, fatto salvo il caso in cui la sostituzione o la modifica di tali atti giuridici comporti una modifica

Coerentemente con la strategia digitale della commissione rivolta alla promozione di una tecnologia umanocentrica, al servizio della persona, la proposta rappresenta un pilastro fondamentale dell'indirizzo politico teso ad assicurare uno sviluppo dell'A.I. rispettoso dei diritti umani. Sui presupposti dell'art. 114 del TFUE la proposta contempla l'applicazione di misure rivolte all'instaurazione del mercato unico digitale, ma, considerato lo scatto in avanti di alcuni stati membri nell'adozione di misure nazionali volte ad assicurare lo sviluppo di un'intelligenza artificiale sicura e rispettosa dei diritti fondamentali, è possibile l'emersione di frammentazione e notevoli vulnerabilità sulla certezza del diritto. Come chiarito anche nella proposta, dal momento che la stessa contiene alcune regole specifiche sul trattamento di dati personali, nello specifico, «restrizioni sull'utilizzo di sistemi di A.I. per l'identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto», è opportuno fare riferimento anche all'art 16 del TFUE[35]. Per rimediare alla potenziale frammentazione, nella proposta è indicato come l'adozione di un regolamento unico possa agevolare «parità di condizioni e tutelerà tutte le persone, rafforzando allo stesso tempo la competitività e la base industriale dell'Europa nel settore dell'A.I.». Il principio di proporzionalità, inteso come pilastro dell'architettura normativa, dovrebbe essere garantito da un approccio duro al rischio alto con imposizione di oneri solo in caso di potenziale vulnus per la sicurezza e la tutela delle libertà fondamentali. Come indicato nella relazione alla proposta «i requisiti di qualità elevata dei dati, documentazione e tracciabilità, trasparenza, sorveglianza umana, precisione e robustezza sono strettamente necessari per attenuare i rischi per i diritti fondamentali e la sicurezza posti dall'A.I. e che non sono oggetto di altri quadri giuridici in vigore». Per tale motivo, l'impiego dell'atto regolamentare, caratterizzato dall'applicazione uniforme e vincolante delle regole in tutti gli stati è ritenuto lo strumento giuridico più idoneo ad assicurare il raggiungimento degli

significativa della progettazione o della finalità prevista del sistema di *A.I.* o dei sistemi di *A.I.* interessati».

[35] La proposta chiarisce anche questo aspetto indicando al paragrafo 2.2 della relazione dedicato alla sussidiarietà indicando come «la natura dell'*A.I.*, che si basa spesso su *set* di dati di grandi dimensioni e varietà che possono essere integrati in qualsiasi prodotto o servizio che circola liberamente nel mercato interno, implica che gli obiettivi della presente proposta non possano essere conseguiti in maniera efficace dai singoli Stati membri. Il formarsi di un mosaico di regole nazionali potenzialmente divergenti potrebbe inoltre ostacolare la circolazione senza soluzione di continuità di prodotti e servizi collegati ai sistemi di *A.I.* in tutta l'UE e potrebbe dimostrarsi inefficace nel garantire la sicurezza e la protezione dei diritti fondamentali e dei valori dell'Unione nei diversi Stati membri. Gli approcci nazionali destinati ad affrontare tali problemi creerebbero soltanto incertezza e ostacoli ulteriori e rallenterebbero l'adozione dell'*A.I.* da parte del mercato».

obiettivi[36].

5. Il controllo umano (*governance* e *due diligence*).

Uno degli aspetti più interessanti del nuovo impianto normativo sarà l'obbligo per i futuri fornitori dei sistemi di A.I. di informazione in caso di "incidenti gravi o malfunzionamenti che costituiscono una violazione degli obblighi in materia di diritti fondamentali" non appena ne vengono a conoscenza, nonché in merito a qualsiasi richiamo o ritiro di sistemi di A.I. dal mercato. «Le autorità nazionali competenti indagheranno quindi sugli incidenti o sui malfunzionamenti, raccoglieranno tutte le informazioni necessarie e le trasmetteranno periodicamente alla Commissione con metadati adeguati». Al titolo II viene elencata una serie di sistemi A.I. ritenuti vietati, differenziandoli in base ad usi con rischi ritenuti a) inaccettabili, b) alti, c) bassi, d) minimi. Più nello specifico, «i divieti riguardano pratiche che presentano un elevato potenziale in termini di manipolazione delle persone attraverso tecniche subliminali, senza che tali persone ne siano consapevoli, oppure di sfruttamento delle vulnerabilità di specifici "gruppi vulnerabili", quali i minori o le persone con disabilità, al fine di distorcerne materialmente il comportamento in maniera tale da provocare loro od a un'altra persona un danno psicologico o fisico. Altre pratiche manipolative o di sfruttamento che interessano gli adulti che potrebbero essere facilitate dai sistemi di A.I. potrebbero essere soggette alla normativa vigente in materia di protezione dei dati, tutela dei consumatori e servizi digitali, che garantisce che le persone fisiche siano adeguatamente informate e dispongano della libera scelta di non essere soggette a "profilazione" o ad altre pratiche che potrebbero influire sul loro comportamento. La proposta vieta altresì l'attribuzione di un "punteggio sociale" basato sull'A.I. per finalità generali da parte di autorità pubbliche. È infine vietato anche il ricorso a sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto alla "criminalità", fatta salva l'applicazione di talune eccezioni limitate»[37]. I sistemi ritenuti ad "alto rischio", ovvero, quelli in cui sono direttamente coinvolti salute e libertà fondamentali, sono consentiti solo con il rispetto di taluni requisiti ed una valutazione di conformità preventiva. L'inserimento di un sistema nella categoria ad alto rischio dipende essenzialmente da modalità di funzionamento e finalità specifiche cui è rivolto. Il capo 1 del titolo III individua due categorie principali:

a) · i sistemi di A.I. destinati ad essere utilizzati come componenti di sicurezza

[36] Coerentemente all'articolo 288 TFUE.

[37] La relazione alla proposta al paragrafo 5.2.2 indica che «l'elenco delle pratiche vietate di cui al titolo II comprende tutti i sistemi di *A.I.* il cui uso è considerato inaccettabile in quanto contrario ai valori dell'Unione, ad esempio perché viola i diritti fondamentali».

di prodotti soggetti a valutazione della conformità ex ante da parte di terzi;
b) · altri sistemi di A.I. indipendenti che presentano implicazioni principalmente in relazione ai diritti fondamentali esplicitamente elencati nell'allegato III.
La commissione può ampliare tale elenco sulla base delle evoluzioni e del progresso nel settore specifico[38]. Ulteriori obblighi specifici di trasparenza, invece, sono previsti per sistemi che: a) interagiscono con gli esseri umani; b) sono utilizzati per rilevare emozioni o stabilire un'associazione con categorie ("sociali") sulla base di dati biometrici; c) generano o manipolano contenuti ("deep fake"). In caso di interazione con sistemi di intelligenza artificiale in grado di valutare emozioni in modo automatizzato, le persone hanno diritto ad essere informate, in modo tale che si possa effettuare una scelta informata con la possibilità di non sottoporsi a tale meccanismo. Ai fini di una governance adeguata ed uniforme sull'intero territorio dell'Unione, la proposta istituisce un «comitato europeo per l'intelligenza artificiale (il "comitato"), costituito da rappresentanti degli Stati membri e della Commissione. Tale comitato faciliterà un'attuazione agevole, efficace e armonizzata del presente regolamento contribuendo all'efficacia della cooperazione tra le autorità nazionali di controllo e la Commissione nonché fornendo consulenza e competenze alla Commissione. Raccoglierà e condividerà inoltre le migliori pratiche tra gli Stati membri». Pur non essendo ancora esistente una definizione chiara ed univoca di cosa sia l'intelligenza artificiale, è pur vero che, fin dalla nascita del termine negli anni sessanta ad opera del matematico McCarthy, i processi computazionali riconducibili al fenomeno di apprendimento automatizzato hanno subito l'attenzione della comunità di studiosi intenzionati a carpirne potenzialità e pericoli. Roger

[38] Il capo 2 definisce «i requisiti giuridici per i sistemi di *A.I.* ad alto rischio in relazione a dati e governance dei dati, documentazione e conservazione delle registrazioni, trasparenza e fornitura di informazioni agli utenti, sorveglianza umana, robustezza, accuratezza e sicurezza...tali requisiti sono altresì in gran parte coerenti con altre raccomandazioni e altri principi internazionali, circostanza questa che assicura che il quadro dell'*A.I.* proposto sia compatibile con quelli adottati dai *partner* commerciali internazionali dell'UE. Le soluzioni tecniche precise atte a conseguire la conformità a tali requisiti possono essere previste mediante norme o altre specifiche tecniche o altrimenti essere sviluppate in conformità alle conoscenze ingegneristiche o scientifiche generali, a discrezione del fornitore del sistema di *A.I.*». Il capo 3 definisce «una serie chiara di obblighi orizzontali per i fornitori di sistemi di *A.I.* ad alto rischio. Obblighi proporzionati sono imposti anche a utenti e altri partecipanti lungo la catena del valore dell'*A.I.* (ad esempio importatori, distributori, rappresentanti autorizzati)». Il capo 4 definisce «il quadro per gli organismi notificati che saranno coinvolti come terze parti indipendenti nelle procedure di valutazione della conformità, mentre il capo 5 spiega in dettaglio le procedure di valutazione della conformità da seguire per ciascun tipo di sistema di *A.I.* ad alto rischio».

Schank, uno dei fondatori del metodo computazionale, tentò già negli anni ottanta di identificarne i caratteri essenziali nella coopresenza di 5 elementi distinti: capacità comunicativa; conoscenza di sé; conoscenza del mondo esterno; comportamento teso ad un fine; alto grado di creatività. La conseguenza logica di tale approccio risiede nell'impossibilità di immaginare tali sistemi solo come surrogati umanoidi nelle manifestazioni estetiche, soprattutto nelle fasi iniziali. Inoltre, aspetto più rilevante, poco o nulla dell'intelligenza artificiale è direttamente riconducibile alle ancora inesplorate complessità dei meccanismi della mente umana, soprattutto per ciò che attiene agli ambiti emotivi e della psiche relazionale[39]. L'intelligenza artificiale mira, più che altro, al raggiungimento di una scelta razionale, percependo, mediante innumerevoli sensori ed elaborando moli impressionanti di dati in poco tempo, l'ambiente circostante, decidendo, infine, la scelta migliore da attuare. Da ciò ne deriva che bisogna sempre considerare alcuni aspetti:
«-i "sensori" potrebbero essere fotocamere, microfoni, una tastiera, un sito Internet o altri sistemi di immissione dati, nonché sensori di quantità fisiche (ad esempio, sensori di temperatura, di pressione, di distanza, di forza/coppia o sensori tattili); -i "dati" acquisiti tramite i sensori sono dati digitali, di cui oggi vi è un'immensa disponibilità; e a proposito dei dati va fin da subito sottolineato che la qualità del risultato finale dipende, in larga misura, proprio dalla correttezza logica e dalla completezza dei dati raccolti; per contro, se i dati utilizzati per alimentare o addestrare il sistema di A.I. sono distorti, nel senso che non sono sufficientemente equilibrati o inclusivi, il sistema non sarà in grado di generalizzare in maniera corretta e potrebbe adottare decisioni inique che possono favorire alcuni gruppi rispetto ad altri.; -il "ragionamento" o l'elaborazione delle informazioni è un processo operato attraverso un algoritmo che acquisisce come input i suddetti dati per poi proporre un'azione da intraprendere alla luce dell'obiettivo da raggiungere; -infine, il sistema di A.I. esegue l'azione prescelta tramite gli "attuatori" a sua disposizione, che possono essere sia software, sia elementi fisici (ad esempio, bracci articolati, ruote automatiche), quest'ultimi capaci di intervenire, modificandolo, sull'ambiente circostante»[40]. Da ciò si evince come in realtà si tratta più propriamente di sistemi c.d. di "machine learning", in cui un software non fa altro che imparare costantemente dall'ambiente esterno,

[39] Gli esperti di *A.I.* preferiscono parlare di razionalità, laddove per "*razionalità*" si intende la «capacità di scegliere la migliore azione da intraprendere per conseguire un determinato obiettivo alla luce di alcuni criteri di ottimizzazione delle risorse a disposizione». Così, S. RUSSELL, P. NORVIG, in *Artificial intelligence: A Modern Approach, Prentice Hall,* 3ª edizione, 2009, pp. 36 ss.
[40] F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo, rivista ,* 29 settembre 2019.

modificando e adattando il proprio comportamento in funzione dei risultati ottenuti con l'elaborazione veloce ed automatizzata dei dati secondo parametri di riferimento preimpostati. Al momento, non esiste ancora una vera e propria tecnologia di A.I., quella che gli studiosi più esperti chiamerebbero "singolarità creativa"[41]. I gruppo indipendente di 52 esperti sull'intelligenza artificiale ha elaborato una definizione di Intelligenza artificiale[42]: «sistemi software (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 A.I. 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'A.I. comprende diversi approcci e diverse tecniche, come l'apprendimento automatico (di cui l'apprendimento profondo e l'apprendimento per rinforzo sono esempi specifici), 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 cyberfisici)».

6. *Law enforcment* ed attività investigative.

L' impiego di tali sistemi nei processi di "*Law enforcement*" è già realtà. Come si evince dal documento di presentazione al convegno annuale degli esperti di polizia, organizzato dall'OSCE: «nei loro sforzi per aumentare l'efficienza e l'efficacia e per stare al passo con le innovazioni tecnologiche, le autorità e le

[41] Una recente Comunicazione del 2018 elaborata dalla Commissione europea, intitolata *"Artificial Intelligence for Europe"* fornisce la seguente definizione di *A.I.*: «l'intelligenza artificiale (*A.I.*) 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'*A.I.* 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'*A.I.* in dispositivi *hardware* (ad esempio, in *robot* avanzati, auto a guida autonoma, droni o applicazioni dell'Internet delle Cose)».

[42] Si veda anche il documento "*Una definizione di A.I.: principali capacità e discipline scientifiche"*, elaborato dal *Gruppo Indipendente di 52 esperti* nominato dalla Commissione europea per svolgere a suo favore funzioni di consulenza sull'intelligenza artificiale.

agenzie di law enforcement di tutto il mondo stanno esplorando sempre più i potenziali dell'A.I. per il loro lavoro. La crescente quantità di dati ottenuti e archiviati dalla polizia ha anche richiesto metodi e strumenti più sofisticati per la loro gestione e analisi, per l'identificazione di modelli (pattern), la previsione dei rischi e lo sviluppo di strategie per allocare le risorse umane e finanziarie dove sono maggiormente necessarie. Anche se l'uso dell'A.I. nel lavoro delle forze dell'ordine è un argomento relativamente nuovo, alcuni strumenti basati sull'intelligenza artificiale sono già stati testati e sono persino attivamente utilizzati dai servizi di polizia di diversi Paesi del mondo. Questi includono software di analisi di video e immagini, sistemi di riconoscimento facciale, di identificazione biometrica, droni autonomi e altri robot e strumenti di analisi predittiva per prevedere le "zone calde" del crimine o anche per identificare potenziali criminali futuri, in particolare i criminali ad elevata pericolosità»[43]. Alcuni dei risvolti derivanti dall'impiego di tale tecnologia nell'attività di contrasto al fenomeno criminale sono stati messi in luce dallo studioso Giuseppe italiano. In un suo saggio ci ricorda alcuni episodi specifici in cui l'impiego della tecnologia di A.I. avrebbe potuto giocare un ruolo determinante. Più nello specifico, «è famoso il caso di Umar Farouk Abdulmutallab, noto anche come il "terrorista delle mutande" (Underwear Bomber), che è riuscito a imbarcarsi sul volo Amsterdam-Detroit nel giorno di Natale 2009, con dell'esplosivo cucito all'interno della biancheria intima che indossava, e che ha cercato di farsi esplodere durante il volo. Per una fortunata coincidenza (l'intervento di alcuni passeggeri insospettitisi), l'attacco terroristico non è andato a buon fine... l'intelligence aveva dati e informazioni a sufficienza per valutare il grado di pericolosità del terrorista e aveva anche elementi sufficienti per inserirlo nella black list, così da negargli la possibilità di imbarco su voli diretti negli Stati Uniti». Come ci ricorda Italiano, in quel caso specifico, l'attività di intelligence «non è riuscita a connettere le molteplici informazioni, provenienti da varie fonti, che erano a sua disposizione. Come dire, nel patrimonio informativo c'erano tutti i dati necessari, ma è semplicemente mancata l'utilizzazione di un buon algoritmo per mettere in correlazione tutti questi dati». Le conclusioni riflessive di Italiano lo portano a rilevare che «sicuramente, e non soltanto in questa circostanza, tecniche di A.I. sono e possono essere impiegate con successo nell'analisi delle informazioni disponibili, delle transazioni, dei file di log, del traffico sulla rete, e di tutte le "impronte" che ogni individuo lascia in rete e nei sistemi digitali, allo scopo di identificare possibili anomalie e attività sospette, o, semplicemente, per comporre in una visione coerente le informazioni provenienti da sorgenti multiple ed eterogenee, ed estrarne conoscenza, in modo tale da prendere in maniera automatica decisioni

[43] *OSCE Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement: An Ally or an Adversary*?, 23-24 September 2019, Vienn.

oppure fornire il supporto a decisori umani, che devono essere in grado di reagire sempre più velocemente agli stimoli esterni»[44]. Dal momento che il controllo umano assurge a fattore determinante nella gestione delle tecnologie di intelligenza artificiale, è opportuno chiedersi se esso debba limitarsi alla scelta di obiettivi e monitoraggio, oppure debba spingersi sino al punto da compromettere le prestazioni del sistema. Un quesito di non facile risposta, cui, presumibilmente sarà attribuita una rimodulazione progressiva e dinamica in funzione del target di riferimento. Molto dipenderà dal gradiente di rischio dell'attività svolta dal sistema o dal livello di interesse rivolto al raggiungimento del risultato aspettato, soprattutto, alla luce delle aspettative di operatori ed opinione pubblica. Da ciò ne consegue che del tutto diverso sarà l'impiego dei sistemi di A.I., quali supporto nelle attività di contrasto a fenomeni criminali di particolare rilievo, nei settori della lotta al terrorismo internazionale e criminalità organizzata, rispetto ad usi finalizzati al contrasto di criminalità semplice o cosiddetta bagattellare. Paradossalmente, è proprio in tali tipologie di attività che il coefficiente di tolleranza all'implementazione di tali sistemi sarà inversamente proporzionale al concreto pericolo di abuso nel loro impiego alla luce dei più sensibili aspetti legati alle principali libertà fondamentali[45]. Tali tipologie di sistemi algoritmici dovranno essere soggette ad un monitoraggio costante, da un punto di vista non solo tecnico, ma, anche, giuridico, in modo da poter implementare un preciso assetto normativo che ne disciplini il legittimo impiego, nel pieno rispetto dei diritti umani.

7. Il contesto penale

La possibilità concreta che in futuro possano verificarsi situazioni di impiego dei sistemi di intelligenza artificiale nelle decisione giudiziarie, nel sistema penale, ha allarmato il consiglio europeo al punto da spingerlo ad adottare la Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti[46], un documento elaborato dalla

[44] G. F. ITALIANO, *Intelligenza artificiale: passato, presente, futuro*, cit., p. 222.

[45] Si pensi, a titolo di esempio, al tema legato alla privacy, o più propriamente, alla tutela del trattamento dei dati personali, in considerazione della gran mole di dati che questi sistemi di A.I. (a volte equipaggiati per l'appunto di sensori e telecamere avanzate) possono acquisire in relazione alla vita, anche privata, dei cittadini coinvolti dal loro utilizzo nei più svariati ambiti. Situazioni in cui i dati oggetto del processo di elaborazione dell'algoritmo potrebbero essere in varie forme manipolati, sottratti o deformati, anche con notevole pregiudizio per le persone cui essi fanno rifermento. Al riguardo, si veda sul tema F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo, rivista online,* 29 settembre 2019.

[46] Si veda S. QUATTROCOLO, I*ntelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*,

Commissione per l'efficacia della giustizia (CEPEJ), il 4 dicembre 2018 nella consapevolezza di un incremento dell'impiego dell'intelligenza artificiale nelle "moderne società". Nella carta vengono specificate delle linee guida a cui dovranno «attenersi i soggetti pubblici e privati responsabili del progetto e sviluppo degli strumenti e dei servizi della I.A.». Si tratta per l'appunto di principi riguardanti il "rispetto dei diritti fondamentali", la "non discriminazione", la garanzia di "qualità" e "sicurezza" dei sistemi, il rispetto dei principi di "trasparenza", "imparzialità" e "correttezza" nel loro impiego ed in particolar modo la garanzia del costante "controllo umano"[47]. Come evidenziato anche dal documento esplicativo allegato alla carta, al 2018, anno in cui venne elaborato, «l'uso di algoritmi di intelligenza artificiale nei sistemi giudiziari europei rimane principalmente un'iniziativa commerciale del settore privato, rivolta a compagnie assicurative, uffici e studi legali, avvocati e privati». Nel documento è evidenziato come tali algoritmi siano meritevoli di essere presi in considerazione nell'ambito della giustizia civile, commerciale e amministrativa «al fine di una risoluzione precontenziosa online delle controversie, purché un ricorso successivo al giudice rimanga possibile». Per quanto riguarda, invece, i procedimenti penali, nel documento è evidenziato che «anche se non sono specificamente progettati per essere discriminatori, l'uso di algoritmi basati sull'A.I. [...] ha mostrato il rischio di favorire la rinascita di teorie deterministiche a scapito delle teorie dell'individualizzazione della pena»[48]. Il rischio, concreto, di incorrere in fenomeni discriminatori o, peggio ancora, pericolosi automatismi durante l'impiego di sistemi algoritmici in ambito penale, deriverebbe, essenzialmente, dal carattere accusatorio del processo nel contesto europeo in cui il giudizio è basato, principalmente, su valutazione di prove dichiarative in merito a cui, un qualsiasi impianto tecnologico di valutazione, avrebbe serie difficoltà a considerarne gli aspetti critici dovuti a menzogne, reticente od atteggiamenti emotivi. Alla luce di quanto stabilito dall'art. 192, comma 2 c.p.p. sui requisiti di chiarezza, precisione e concordanza degli elementi di prova, sarebbe molto difficile per un sistema di intelligenza artificiale, ancorché in un processo di tipo indiziario, poter valute il grado di coerenza logica delle prove sulla base dei criteri richiesti. Del tutto impossibile per un sistema di intelligenza artificiale è l'applicazione della regola di giudizio stabilità dall'art. 533, comma del c.p.p. su "oltre ogni ragionevole dubbio", in quanto, al momento, i software in questione hanno capacità di elaborazione solo con logiche binarie (bianco-nero; vero-falso; sì-no) o probabilistiche

in *www.lalegislazionepenale.eu*, 18 dicembre 2018.

[47] Il cui fine è «rivolto a precludere un approccio deterministico» e ad «assicurare che gli utilizzatori agiscano come soggetti informati ed esercitino il controllo delle scelte effettuate». Virgolettati estratti dalla carta.

[48] Rispettivamente pag. 16, 41 e 48 del documento.

in %[49]. Il quesito fondamentale cui rispondere riguarda la possibilità che un soggetto con determinate caratteristiche possa commettere un reato in futuro. Come sostenuto da Basile nel suo recente saggio del 2019 su diritto penale e uomo dal titolo Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine, «Si tratta di un quesito la cui risposta è necessaria, tra l'altro, quando si tratta di applicare una misura di sicurezza, una misura cautelare o una misura di prevenzione, o anche per concedere la sospensione condizionale di una pena o l'affidamento in prova al servizio sociale. Ebbene, a tale fondamentale quesito oggi i nostri giudici forniscono risposte per lo più intuitive, affidate esclusivamente alla loro esperienza personale e al loro buon senso, oppure, quando consentito dalla legge, basate su valutazioni cliniche di periti, mentre, in futuro, (e già nel presente di altri ordinamenti giuridici), siffatte valutazioni prognostiche della pericolosità criminale potrebbero essere affidate a specifici algoritmi (risk assessment tools, o algoritmi predittivi), capaci di attingere e rielaborare quantità enormi di dati al fine di far emergere relazioni, coincidenze, correlazioni, che consentano di profilare una persona e prevederne i successivi comportamenti, anche di rilevanza penale». Il sistema di valutazione del rischio criminale, nella maggior parte degli algoritmi predittivi, è basato essenzialmente sul c.d. "criterio attuariale" in cui la determinazione del rischio avviene sulla base di riscontri oggettivi tesi ad integrare o, più precisamente, a sostituire il ruolo del giudice nella sua attività di gestione del caso sottoposto al suo esame. Si tratta in sostanza di una determinazione del rischio "evidence-based", data dall'individuazione di alcuni "fattori predittivi" ritenuti integranti il comportamento criminale quali: «età, sesso, origine etnica, livello di scolarizzazione, situazione familiare e lavorativa, posizione sociale, precedenti penali, precedenti esperienze carcerarie, luoghi e le persone frequentati, presenza di autori di reato nella cerchia familiare o nella rete di conoscenze, luogo di residenza, difficoltà di regolazione della rabbia e aggressività, dis-controllo degli impulsi, storia di precedente violenza agita, storia di ospedalizzazione, pensiero pro-criminale, variabili contestuali (la mancanza di sostegno familiare e sociale), consumo di sostanze stupefacenti o alcoliche, psicopatie»[50]. Dall' approccio statistico

[49] G. CANZIO, *Il dubbio e la legge*, in *Diritto penale contemporaneo*, 2018, pp. 1 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Unite ed Europa*, in ivi, 2019, pp. 1 ss.; A. NATALE, Introduzione. *Una giustizia (im) prevedibile?*, in *Questione Giustizia*, fasc. 4, 2018, pp. 1 ss.; nello stesso fascicolo, v. pure i contributi di C. COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, e di C. CASTELLI, D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*.

[50] G. ZARA, *Tra il probabile e il certo. La valutazione dei rischi di violenza e di recidiva criminale,* in *Diritto penale contemporaneo*, 20 maggio 2016, con riferimento, in

nella valutazione di tali fattori di rischio, è possibile giungere alla determinazione di "scale" di valore della pericolosità criminale del soggetto sottoposto a giudizio in cui vi può essere una «ulteriore differenziazione basata sul tipo di "popolazione" di riferimento (adulti, minori, maschi, femmine, pazienti psichiatrico-forensi, detenuti o ex-detenuti), alla tipologia di reati contestati ("scale generiche", relative a tutti i reati, e "scale specifiche", relative a singole tipologie di reati), alla "temporalizzazione" del rischio (immediato, a medio o a lungo termine) e, infine, al "contesto applicativo" (comunità civile, istituti di pena, centri di salute mentale, ospedali psichiatrico-giudiziari)[51]». I vari fattori rischio contemplati nella valutazione del rischio possono avere, inoltre, un differente grado di dinamicità. Dal momento che, secondo Gaza, esistono "fattori statici", non modificabili (sesso, origine etnica), "fattori dinamici" stabili, modificabili con la sottoposizione del soggetto ad un trattamento terapeutico e, infine, "fattori di rischio acuti", soggetti a rapidi cambiamenti o associati ad una condizione pregressa che ne agevola la reazione violenta (uso di sostanze stupefacenti)[52.] Da ultimo, ma non per importanza, è da considerare il concetto di c.d. "dose-exposure relationship", il quale, secondo lo stesso Gaza, dipende da «precocità, durata e intensità dell'esposizione a più fattori di rischio che interagiscono in modo cumulativo, equifinale, dinamico, aumentano la probabilità di violenza e manifestazioni criminali»[53.] Chi sostiene la possibilità di un futuro utilizzo degli algoritmi di intelligenza artificiale nelle aule di giustizia muove dall'assunto che, simili sistemi, grazie all'elaborazione di dati ed autoapprendimento, possano agevolare valutazioni di pericolosità più accurate ed esenti dai rischi umani di pregiudizio. Come ha potuto dimostrare anche l'esempio statunitense del sistema COMPAS, innumerevoli sono anche le perplessità relative a tali sistemi, con riguardo ad aspetti di accuratezza e trasparenza degli algoritmi cui si basano le decisioni. Come messo in luce da Giovanni Canzio, non mancano anche enormi problemi etici e deontologici legati al loro impiego in ambito giudiziario: «il dubbio del giudicante in ordine alla propensione dell'imputato a ripetere il delitto non trova più la soluzione in un criterio metodologico di accertamento del fatto e neppure in una puntuale prescrizione della legge, ma viene affidato a un algoritmo di valutazione del rischio, elaborato da un software giudiziario[...]. Considerati i risultati pratici – soprattutto in termini di risparmio – conseguiti dall'impiego del modello matematico-statistico, neppure le cautele e il

particolare, al lavoro di J. P. SINGH et al., *A comparative study of violence risk assessment tools: a systematic review and meta-regression analysis of 8 studies involving 25980 participants*, in Clin Psychol Rev, 31, 2011, pp. 499 ss.

[51] G. ZARA, *tra il probabile e il certo*, cit., pp. 17 ss.

[52] G. ZARA, *Tra il probabile e il certo*, cit., p. 12

[53] G. ZARA, *Tra il probabile e il certo*, cit., p. 14

warning delle Corti e lo scetticismo degli studiosi, quanto al rispetto delle garanzie del due process nella raccolta delle informazioni utili per la valutazione del rischio nel mondo reale e all'eventuale pregiudizio discriminatorio, sono riusciti a frenare l'impetuosa avanzata delle tecniche informatiche di tipo predittivo nel sistema statunitense di giustizia penale. Si è forse agli inizi di uno sconvolgente (e però non auspicabile) mutamento di paradigma della struttura e della funzione della giurisdizione? A fronte della complessità tecnica, dei tempi e dei costi delle faticose operazioni giudiziali ricostruttive del fatto, la postmodernità metterà in crisi l'equità, l'efficacia e le garanzie del modello proprio del razionalismo critico, oppure resterà ben salda e vitale l'arte del giudicare reasonig under uncertainty, seppure by probabilities?»[54.] Nel contesto europeo, almeno per il momento, non si intravede la possibilità di un ingresso degli algoritmi predittivi nelle aule penali, anche alla luce dell'art. 15 della direttiva 95/46/CE, confluito nell'art. 22 del nuovo Regolamento europeo in materia di protezione dei dati personali, entrato in vigore il 25 maggio 2018[55.] A supporto del ragionamento, la Risoluzione del Parlamento europeo sulla robotica del 2017 fa leva esattamente sul principio della "trasparenza", evidenziando la necessità che sia sempre possibile individuare la logica alla base di ogni decisione, cui sia coinvolto l'uso od aiuto dell'intelligenza artificiale, soprattutto, qualora tale decisione possa determinare un impatto significativo sulla vita di una o più persone coinvolte nel giudizio. Andrea Natale, elabora alcune considerazioni sintetiche degne di essere riportate: «(a) il risultato fornito dagli algoritmi predittivi è necessariamente influenzato dalla qualità dei dati che vengono posti come input; ne discende che è indispensabile prevedere meccanismi che assicurino: (a.1) la qualità del dato; (a.2) l'indipendenza della fonte da cui provengono i dati; (a.3) l'indipendenza dell'autorità che raccoglie i dati; (a.4) l'accessibilità a tutti dei dati posti come input dell'algoritmo; (b) è necessario scongiurare il rischio che l'algoritmo possa avere un esito discriminatorio fondato su dati personali sensibili, tra cui la razza e l'estrazione sociale [...]; (c) la verificabilità o meno della struttura dell'algoritmo; un algoritmo ha una sua struttura che non è neutra [...]; nel concepire l'architettura di un algoritmo, il programmatore fa delle scelte che, necessariamente, influenzano il risultato dell'operazione computazionale; il programmatore può fare degli errori di progettazione; un algoritmo la cui

[54] G. CANZIO, *Il dubbio e la legge*, in *Diritto penale contemporaneo*, 20 luglio 2018, pp. 3 s. Sul classico ragionamento giudiziario *"by probabilities*", v. lo stesso G. CANZIO, L*a motivazione della sentenza e la prova scientifica: "reasoning by probabilities*", in G. CANZIO, L. LUPARIA (a cura di), *Prova scientifica e processo penale,* Cedam, 2018, pp. 3 ss.

[55] L'articolo, infatti, stabilisce il divieto di giudizi automatizzati basati esclusivamente sulla personalità del soggetto.

struttura sia protetta da diritti di proprietà intellettuale e non open source è sottratto alla possibilità di controllo, verifica e confutazione da parte della parte processuale e, più in generale, della comunità degli utenti; ciò comporta non pochi problemi, tanto sotto il profilo della validazione dell'affidabilità scientifica del risultato che l'algoritmo restituisce, quanto sotto il profilo del diritto di difesa; si ritiene, pertanto, indispensabile che – laddove si voglia davvero fare un uso processuale di algoritmi predittivi da parte del sistema giudiziario (che è un sistema per sua natura pubblico) – nessun segreto possa essere posto sull'architettura degli algoritmi e dei dati che lo alimentano; si deve poi elaborare un meccanismo che assicuri anche l'indipendenza di chi ha elaborato l'algoritmo (che senso ha costituzionalizzare l'indipendenza del giudice e la sua soggezione solo alla legge se non si coltiva analoga pretesa a chi elabora uno strumento decisorio di simile portata?); (d) l'algoritmo – anche ove usato non come strumento decisorio esclusivo, ma come mero supporto alla decisione del giudice – richiede formazione; è dunque indispensabile formare il personale giudiziario che potrebbe doversene avvalere; (e) l'algoritmo predittivo – muovendo da una elaborazione della giurisprudenza e dei casi precedenti – può indicare non il "risultato" esatto di una certa controversia, ma il suo possibile esito, evidenziando quali siano le linee giurisprudenziali prevalenti e quali gli esiti concreti che si sono dati in casi simili; ciò, però, comporta più di un rischio: (e.1) l'algoritmo non è in grado di "riconoscere" che quello a lui sottoposto non è un caso simile; vi sono delle singolarità che un decisore umano, forse rileverebbe e che lo porterebbero ad operare un distinguishing; l'algoritmo non è progettato per prevedere questa evoluzione; (e.2) per la stessa ragione, l'algoritmo può favorire quello che Garapon [...] chiama come "effetto moutonnier" (effetto pecora nel gregge): è concreto, in altri termini, il rischio di indurre il giudice pigro ad adagiarsi sulla proposta dell'algoritmo senza assumere su di sé l'autentica responsabilità del giudizio che egli emette; (e.3) per la stessa ragione, l'uso di algoritmi può favorire una cristallizzazione della giurisprudenza, rendendola meno sensibile ai cambiamenti sociali (e, di fatto, rendendoli meno probabili)»[56]. Infine, come precedentemente accennato, le stesse perplessità sono stata sollevate nella Carta etica elaborata dal CEPEJ «l'uso di algoritmi in materia penale, al fine di profilare le persone, è stato criticato dalle ONG a causa dei limiti della metodologia utilizzata, del loro approccio meramente statistico, il quale avrebbe effetti discriminatori e deterministici, sicché esso andrebbe sostituito da un altro approccio che risulti più rispettoso delle norme europee in materia di sanzioni penali e che salvaguardi le chances di riabilitazione e reintegrazione del singolo individuo. Se i sistemi algoritmici riescono a migliorare la raccolta di informazioni per

[56] NATALE, Introduzione. *Una giustizia (im)prevedibile?*, in *Questione Giustizia*, fasc. 4, 2018, pp. 3 ss

valutazioni inerenti la probation, per esempio, e a rendere possibile che le informazioni pertinenti siano raccolte più rapidamente per la successiva elaborazione umana, allora si tratterebbe sicuramente di un progresso (in particolare nei procedimenti sommari). Qualsiasi altro uso è esposto a pregiudizi destinati ad entrare in conflitto con alcuni principi fondamentali, nazionali e sovranazionali»[57].

8. Conclusioni.

In conclusione, le tecnologie di intelligenza artificiale giocheranno, probabilmente, un ruolo determinante e sempre più pervasivo nel contesto di attività sociali proiettate alla massimizzazione dell'efficienza funzionalistica. Allo stesso tempo, progressivamente con l'evoluzione tecnologica, la possibile implementazione di tali sistemi, anche in ambito giudiziario ed in particolar modo nel settore penale, potrebbe rappresentare un'opportunità per gli innumerevoli e lenti ingranaggi della giurisdizione ma, presumibilmente, un profondo vulnus ai principi fondamentali del giusto processo, qualora dovesse evolvere, quale nuovo strumento di supporto ad attività investigative o di giudizio, nel complesso processo di valutazione del giudice nel caso concreto, senza un adeguato controllo umano. Le esperienze fin qui maturate, frutto soprattutto di applicazioni parziali ed in sistemi giuridici di matrice anglosassone (sistema COMPAS), hanno già posto in evidenza i possibili rischi derivanti da un poco attento impiego delle più evolute tecnologie di A.I., soprattutto alla luce del potenziale difetto originario di impostazione dell'algoritmo alla base del funzionamento della macchina, la quale potrebbe risentire in modo negativo sia della scarsa qualità dei dati integrati nel giudizio di valutazione, sia dei diversi bias, anche inconsapevoli, dei suoi programmatori. La recente proposta di regolamentazione europea cerca di porre rimedio in forma preventiva alle più significative criticità derivanti dal potenziale futuro impiego delle tecnologie di A.I.. Particolare attenzione è posta nella gradazione del margine di rischio, laddove se ne prospetta progressivamente il maggiore controllo umano a latere, fino alla totale esclusione d' impiego dei sistemi ad alto rischio, in virtù della loro incidenza sul nucleo fondamentale dei diritti dell'individuo. Strumenti abnormi di controllo massivo quali il credito sociale o tecnologie real time di monitoraggio facciale su larghissima scala non dovrebbero essere implementate sul suolo europeo, almeno nelle intenzioni del legislatore unitario. Sarà, però, interessante verificare la risposta nei singoli ordinamenti a tali vincoli, soggetti a vari grimaldelli di deroga dal contorno assai elastico e poco determinato, nell'ipotesi, oramai consolidata, di contrasto a fenomeni considerati dal particolare allarme sociale, quali terrorismo o criminalità organizzata. Il tema, assai complesso e difficilmente

[57] Carta Etica, cit., pp. 67 ss.

inquadrabile per schemi di semplificazione, coinvolgerà verosimilmente l'interesse e l'attenzione di studiosi ed addetti ai lavori nel futuro prossimo. L'elaborato, lungi dall'essere un quadro esaustivo dello stato dell'arte, anche a fronte degli innumerevoli progressi nel campo, si pone quale obiettivo una prima razionalizzazione delle principali tematiche legate al potenziale impiego delle moderne tecnologie di A.I. nel contesto della "giustizia penale", alla luce delle esperienze già maturate in altri sistemi, nonché dei presupposti normativi interni ed europei. L'impressione maturata durante il processo di ricerca è di un notevole interesse al tema degli addetti ai lavori, ma, allo stesso tempo, una percepibile difficoltà del legislatore nel seguire adeguatamente il velocissimo processo di evoluzione delle tecnologie in questione. Ulteriori esperienze pratiche, alcune preannunciate, altre in fase di studio, potranno fornire migliori risposte alle criticità sollevate da più fronti, ma, nelle more di progressiva consapevolezza al tema delle parti coinvolte (addetti ai lavori, legislatore, opinione pubblica e, soprattutto, sviluppatori), sarà di fondamentale importanza non abbassare la guardia e mantenere alta l'attenzione.