L'ORACOLO ALGORITMICO E LA GIUSTIZIA PENALE: AL BIVIO TRA TECNOLOGIA E TECNOCRAZIA*

dis CRIMEN articoli
15.5.2020

Vittorio Manes

SOMMARIO 1. Giustizia penale e agenti non umani. — 1.1. L'imputazione della responsabilità penale per azioni (autonome) delle macchine. — 2. L'A.I. e il sistema penale… — 2.1. …in sede investigativa. — 2.2. …e in sede giudiziale. — 2.3. La giustizia predittiva penale. — 3. L'esperienza d'Oltreoceano. — 4. Le prospettive di applicazione dell'AI nel sistema penale italiano. — 5. Garanzie fondamentali e problematiche di compatibilità costituzionale. — 5.1. *Machine bias* e principio di eguaglianza. — 5.2. Trasparenza e riserva di legge. — 5.3. Valutazioni statistiche, "diritto penale del fatto" e principio di personalità della responsabilità penale. — 5.4. Giusto processo e tutela del diritto di difesa. — 5.5. Analisi algoritmica dei dati e diritto al silenzio dell'imputato. — 6. Riflessioni conclusive: l'algoritmo come supporto alla decisione giudiziale.

1. Giustizia penale e agenti non umani

Davanti all'avanzare irruento dell'AI e degli algoritmi in campo giuridico, si avverte il disagio di ogni "mutamento di paradigma"[1], se non di una rivoluzione epistemologica[2] che si accompagna ad una vera e propria "frattura antropologica"[3].

Il disagio aumenta – lo si può ben intuire – se si osserva la scena dall'angolatura del sistema della giustizia penale.

---

* Il presente contributo è in corso di pubblicazione nel volume AA.VV., *Intelligenza Artificiale – Il diritto, i diritti, l'etica*, a cura di U. Ruffolo, Giuffré, Milano, 2020.

[1] V. l'autorevole denuncia di G. CANZIO, *Il dubbio e la legge*, in *www.penalecontemporaneo.it*, 20 luglio 2018, 3 s.; altresì ID., *La motivazione della sentenza e la prova scientifica:* "reasoning by probabilities", in AA.VV., *Prova scientifica e processo penale*, a cura di G. Canzio-L. Luparia, Padova, 2018, 3 ss.

[2] Rivoluzione che sta attraversando tutti i settori del diritto: se ne ha un saggio – in una lettura ormai sterminata – leggendo il *focus* curato da E. GABRIELLI e U. RUFFOLO, *Intelligenza artificiale e diritto*, in *Giur. it.*, 2019, 1657 ss.

[3] Di "*rupture anthropologique*" parlano, sin dal sottotitolo del loro saggio, A. GARAPON-J. LASSÉGUE, *Justice digitale. Révolution graphique et rupture antropologique*, Paris, 2018, sul quale v. la recensione di E. FRONZA, *Code is law*. Note a margine del volume di Antoine Garapon e Jean Lasségue, *Justice digitale. Révolution graphique et rupture antropologique*, PUF, Paris, 2018, in *www.penalecontemporaneo.it*, 11 dicembre 2018.

In effetti, il diritto penale è pensato ed edificato sull'uomo, sul rimprovero personale e colpevole, sul grado di responsabilità e di rimproverabilità per una azione *umana*; il processo penale è parimenti affidato ad un giudice *human being*, alla sua capacità di comprensione e di valutazione – secondo una logica valoriale umanamente *fuzzy* –, alla sua ragionevolezza ed equità nell'esercizio di un potere discrezionale, che si esercita nel "crepuscolo del dubbio"; lo strumento di controllo critico di questa discrezionalità vincolata è la *motivazione*, al centro della quale stanno la ragionevolezza e la fondatezza delle argomentazioni pro-poste a suo sostegno[4], o a sostegno delle scelte in punto di commisurazione della pena (art. 133 c.p.); secondo un itinerario di razionalità che, in definitiva, anche nel suo segmento più fluido – l'interpretazione – deve essere sempre sorretto da un "fondamento ermeneutico controllabile"[5].

Insomma, il sistema penale è un sistema personocentrico e personologico, pensato per l'uomo ed affidato al giudizio dell'uomo, come tale *fallibile* ma pur sempre *controllabile* secondo un determinato *iter* argomentativo e i criteri che lo guidano.

Come si sa, questo sistema si è dovuto già misurare con l'impatto – sempre più dispiegato – della *corporate liability*, che ha condotto il diritto penale – faticosamente – ad incriminare e giudicare persone giuridiche, ossia "entità inumane*"* (d.lgs. n. 231 del 2001); e ciò, nonostante fosse difficile identificare, al cospetto di un *corporate crime*, una "condotta" secondo il concetto tradizionale di *azione*, nonostante fosse arduo rintracciare un rimprovero secondo una nozione personalistica di colpevolezza, e nonostante mancasse una "persona" da risocializzare mediante la pena[6].

1.1 - *L'imputazione della responsabilità penale per azioni (autonome) delle macchine*

L'ingresso incalzante dell'intelligenza artificiale – non più solo preconizzato da romanzi distopici e *science fictions* – apre bruscamente un nuovo scenario: il diritto penale, a breve, si dovrà confrontare con macchine in tutto o in parte *self-driving*, ancora una volta "persone senz'anima" per le quali – di fronte all'eventuale causazione

[4] Anche e soprattutto in ordine alla prova, in merito alla quale il giudice deve "[…] dare conto nella motivazione dei risultati acquisiti e dei criteri adottati": art. 192, comma secondo, c.p.p.; così come in ordine alla gravità indiziaria, etc.

[5] Tra le molte decisioni della Corte costituzionale, v. ad es. la sentenza n. 5 del 2004.

[6] Nonostante, insomma, non vi sia alcun corpo da colpire, né alcuna anima da condannare, secondo il titolo del celebre saggio di J. C. COFFEE JR., *No Soul to Damn: No Body to Kick. An Unscandalized Inquiry into the Problem of Corporate Punishment*, in *Michigan Law Review*, vol. 79 (n. 3), 1981, 386 ss.

di un evento di danno o pericolo, colposo o persino "doloso" – molto si potrà e dovrà discutere in punto di imputazione della responsabilità[7].

È un problema che ripropone dilemmi giuridici ma ancor prima problemi e scelte essenziali di politica del diritto, che il legislatore italiano, peraltro, non ha ancora affrontato[8].

Sotto il primo profilo, si tratta di comprendere se gli attuali moduli di attribuzione della responsabilità penale possano adattarsi ad un fatto-reato dove l'*azione* è *opera autonoma* di una macchina guidata da *software* e algoritmi, solo "assistita" dalla presenza umana inerte e solo eventuale; ovvero – caso forse ancor più frequente e complesso – come possano inquadrarsi ipotesi dove la condotta è *opera condivisa* tra agire umano e intelligenza artificiale, via via secondo classificazioni che riconoscono ormai diversi livelli di *driving automation*. E se dunque l'eventuale causazione di un "evento" – colposamente provocato – sia da imputarsi a chi ha *generato* il focolaio di rischio ideando l'algoritmo (specie se *self-learning*)[9], o applicandolo nel programmare il "*robot*" o il PC che guida la macchina; a chi ha *attualizzato* quel rischio producendo e mettendo in commercio il veicolo; ovvero a chi ha concretamente *gestito* quel rischio servendosi della stessa, o magari cooperando con essa (ove sussistano, ad esempio, eventuali profili di *culpa in interagendo*)[10].

Ancora, a questo riguardo, si tratta di valutare se alle diverse costellazioni di casi si adattino schemi di imputazione della responsabilità declinati sul principio della responsabilità personale e colpevole[11], costituzionalmente imposto (art. 27, commi primo e terzo, Cost.), ed orientati alla finalità rieducativa della pena; o se debbano

[7] Per una panoramica sullo scenario aperto dall'interrogativo "*machina delinquere potest*?", v. ad es. F. BASILE, *Intelligenza artificale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. Pen. Uomo*, 2019, 27 ss.

[8] Al momento, all'interno del Consiglio d'Europa, solo un limitato numero di paesi – a quanto consta – ha adottato normative generali per l'utilizzo della guida automatizzata (Austria, Germania, Francia e Svizzera), peraltro ricorrendo alle nozioni tradizionali riferibili ai diversi schemi di responsabilità; mentre altri stati hanno adottato solo norme specifiche concernenti *pilot tests*.

[9] Nella prospettiva civilistica, v. le stimolanti osservazioni di U. RUFFOLO, *Intelligenza artificiale,* machine learning *e responsabilità da algoritmo*, in *Giur. it.*, 2019, 1689 ss.

[10] Con riguardo ai problemi concernenti i veicoli *self-driving*, nella prospettiva civilistica, v. ancora U. RUFFOLO-E. AL MUREDEN, Autonomous vehicles *e responsabilità nel nostro sistema ed in quello statunitense*, in *Giur. it.*, 2019, 1704 ss., rilevando peraltro che "la crescente automazione sposterà via via il peso delle responsabilità dal *driver* al costruttore"; ma anche A. AMIDEI, *Intelligenza Artificiale e* product liability*: sviluppi del diritto dell'Unione europea*, ivi, 1715 ss.

[11] Proprio sulla categoria della colpevolezza in questo mutato contesto si interroga S. BECK, *Digitalisierung und Schuld*, in T. FISCHER, E. HOVEN, *Schuld*, Baden-Baden, 2017, 289 ss.

piuttosto preferirsi schemi di attribuzione della responsabilità – evidentemente eccentrici rispetto ad un rimprovero penale costituzionalmente orientato – basati sulla causazione oggettiva del danno, ed orientati al mero risarcimento, ovvero centrati su nozioni "nuove" come quelle – sono solo ipotesi – di "colpa di programmazione" o "di automazione" che coinvolgano, in prima battuta, l'impresa produttrice della macchina, sul modello appunto della *product liability*[12].

Al tempo stesso, ed ancor prima, si tratta di verificare se il problema dell'imputazione della responsabilità possa essere risolto utilizzando concetti e categorie tradizionali, pur declinate mediante una apposita indicazione legislativa in ordine al terminale ultimo ed all'estensione della responsabilità[13], ovvero – per usare il lessico penalistico – in ordine al titolare della "posizione di garanzia" così come ad estensione e contenuto della stessa; o se la soluzione a tali problemi debba implicare la creazione di nuove nozioni legali, come quella legata alla attribuzione di una "personalità giuridica artificiale"[14] alla macchina in tutto o in parte *self-driving* (una sorta di e-*personhood*), o quella riferibile ad una peculiare forma di cooperazione colposa (*contributory*

[12] Sui quali v. ancora U. RUFFOLO, *Intelligenza artificiale,* machine learning *e responsabilità da algoritmo*, cit., 1694 ss., 1697 ss.; il richiamo al paradigma della responsabilità da prodotto implicherebbe l'apparente paradosso "che i fabbricanti potrebbero essere disincentivati ad orientare la produzione in questa direzione, frenando un progresso tecnologico che dovrebbe avere effetti positivi sulla sicurezza stradale, se è vero che la automatizzazione della guida di veicoli – si ritiene – abbatterà drasticamente il numero di incidenti"; del resto, traslare la responsabilità da rischi da prodotto tecnologicamente evoluto dal produttore all'utente significherebbe disincentivarne l'acquisto: al riguardo, ancora U. RUFFOLO-E. AL MUREDEN, Autonomous vehicles, cit., 1705.

[13] Le soluzioni legislative avanzate sembrano rispecchiare una alternativa, ben espressa dalle opzioni adottate nel modello austriaco e tedesco. Secondo il modello austriaco, "Il guidatore può trasferire alcune funzioni di guida a sistemi (di assistenza alla guida autorizzati), ma resta responsabile in ogni momento di riprendere tutte le funzioni di guida" (art. 3, comma 2, dell'*AutomatFahrV* austriaco); secondo il modello tedesco, "I guidatori possono distogliersi dalla situazione del traffico e cedere il veicolo a meccanismi di guida assistita nella misura in cui essi utilizzano funzioni di guida automatica correttamente e sono pronti a rispondere a una richiesta di riassunzione in ogni istante" (§§ 1° e 1b dello *Strassenverkehrgesetz*).

In questa prospettiva, in particolare, è chiaro che se il guidatore non è richiesto di monitorare il traffico sino a una richiesta di riassunzione della funzione di guida, lo stesso non può più essere ritenuto in concreto "*human in command*" né dunque (penalmente) responsabile di eventuali causazioni lesive occorse sino a quel momento, ove appunto il controllo sulla attività rischiosa era delegato alla macchina *AI driven* legalmente autorizzata.

[14] Aprendo dunque a cascata il dibattito sulla possibilità di punire già l'algoritmo o l'agente intelligente, cfr., *ex multis*, S. GLEß, T. WEIGEND, *Intelligente Agenten und das Strafrecht*, in *ZStW*, 2014, 561 ss.; G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Cham, 2015, 185 ss.; E. HILGENDORF, *Können Roboter schuldhaft handeln?*, in S. BECK, *Jenseits von Mensch und Maschine*, Baden-Baden, 2012, 119 ss.; M. SIMMLER, N. MARKWALDER, *Roboter in der Verantwortung? Zur Neuauflage der Debatte um den funktionalen Schuldbegriff*, in *ZStW*, 2017, 20 ss.; S. ZIEMANN, *Wesen,*

negligence) capace di adattarsi alle interazioni uomo-macchina ed allo specifico "rischio di interconnessione" (*interconnectivity risk*)[15].

Sotto il secondo profilo, ed appunto a monte, si tratta di compiere una ulteriore scelta essenziale – e pregiudiziale – di politica legislativa: se cioè, per esemplificare con il caso attualmente più discusso, riferibile appunto al veicolo *self-driving* immesso nella circolazione stradale ordinaria – una volta certificata l'affidabilità dell'*autonomous vehicle*, autorizzata l'immissione dello stesso sul mercato e consentita quindi la circolazione, l'eventuale causazione di eventi dannosi o pericolosi sia da ascriversi all'area del "rischio consentito" (*Socially Accepted Risk*) ovvero sia da ascriversi (al produttore, come accennato, o) a chi concretamente si è assunto quel rischio, mettendosi alla guida.

Come si vede, un (ulteriore) problema di allocazione del rischio, oggetto di precise scelte politiche, aperte anche a valutazioni di analisi economica del diritto; ed oggetto di possibili discrepanze tra ordinamenti, con conseguente difficoltà di inquadramento dei casi *cross-border*[16].

Non è un caso che l'attenzione su questi problemi sia altissima, anche in seno alle organizzazioni sovranazionali, primo fra tutti il Consiglio d'Europa: l'*European Committee on Crime Problems*, di recente, ha istituito un *Group of Experts on Artifical Intelligence and Criminal Law* espressamente deputato a discutere le varie soluzioni adottate nei diversi Stati membri, le iniziative legislative domestiche intraprese e in corso di studio e, soprattutto, le iniziative da affidare ad eventuali strumenti di armonizzazione sovranazionale, più o meno vincolanti per gli Stati.

2. L'A.I. e il sistema penale…

Ma al di là di questo primo, notevole versante problematico, l'intelligenza artificiale e l'utilizzo degli algoritmi aspira a penetrare alle radici del sistema, toccando i

Wesen, seid's gewesen? Zur Diskussion über ein Strafrecht für Maschinen, in E. HILGENDORF, J.-P. GÜNTHER, *Robotik und Gesetzgebung*, Baden-Baden, 2013, 183 ss.

[15] Così, ad esempio, se una *AI driven car* è alimentata da dati scorretti relativi alla mappa, che indicano un limite di velocità superiore a quello realmente vigente, e, parallelamente, la funzionalità del sensore è limitata e non è in grado di interpretare correttamente la segnaletica stradale.

[16] O ai casi in cui la *AI driven car*, autorizzata in un ordinamento ove si riconosce spazio al "rischio consentito", investa un cittadino di altro Stato, dove tale rischio non viene riconosciuto e conseguentemente si apra un procedimento penale per accertare la responsabilità (penale) di chi ha cagionato l'evento.

più diversi ambiti – dal *policing* al *profiling* al *sentencing*, in prospettiva sia *ante delictum* che *post delictum* –, e sfida apertamente il "fattore umano" che informa di sé il sistema penale: prospettando come alternativa un "sistema oracolare *legal-tech*".

Essa, infatti, ambisce a migliorare le prestazioni del sistema preventivo e repressivo operando a diversi livelli, promettendo un eccezionale *improvement* di efficacia ed efficienza, o persino il definitivo coronamento dei suoi obiettivi (la tutela dei beni giuridici); ma al contempo prospetta tensioni con i diritti fondamentali ed autentiche sfide etiche – che del resto hanno già sollecitato diverse risposte istituzionali, ancora, sul fronte sovranazionale[17] – lasciando persino temere – ben oltre, se si vuole, la dicotomia antica tra *crime control* e *due process*[18] – la scomparsa del diritto penale e l'eclissi dei suoi principi fondamentali[19].

2.1 - *…in sede investigativa*

In sede *investigativa* e di *polizia*, promette di migliorare l'efficientamento delle risorse di *law enforcement* migliorando le attività di *policing* (c.d. *predictive policing*)[20] e le tecniche di *profiling* (attraverso sistemi di riconoscimento facciale, di identificazione biometrica, etc.).

Da un lato, tali programmi consentono di "mappare" il rischio criminale e provvedere ad una razionale allocazione delle risorse per neutralizzare la commissione di prevedibili reati e così ridurre la vittimizzazione (così, ad esempio, il programma *Keycrime*, che trae origine da esperienze investigative presso la Questura di Milano,

[17] Nel dicembre 2018 la *Commissione per l'efficienza della giustizia del Consiglio d'Europa* (CEPEJ) ha adottato la *Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia*, documento che contiene cinque principi generali: rispetto dei diritti fondamentali; non discriminazione; qualità e sicurezza dei dati; trasparenza, imparzialità e *fairness*; possibilità di controllo da parte dell'utente: al riguardo, v. S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *www.lalegislazionepenale.it*, 18 dicembre 2018.

[18] Al riguardo, ora, D. NEGRI, *Modelli e concezioni*, in A. CAMON-C. CESARI-M. DANIELE-M.L. DI BITONTO-D. NEGRI-P. PAULESU, *Fondamenti di procedura penale*, Padova, 2019, 13 ss.

[19] Di fronte all'incalzare dell'AI, sulla fine del diritto penale – per esaurimento dei suoi scopi o, viceversa, per tramonto dei suoi principi – si interroga C. BURCHARDT, *L'intelligenza artificiale come fine del diritto penale?*, dattiloscritto in corso di pubblicazione.

[20] Sul punto, rispetto all'esperienza statunitense, W.S. ISAAC, *Hope, Hype, and Fear: The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice*, in *Ohio St. J. Crim. L.*, 2018, 543 ss.; nel contesto italiano, non senza accenti critici, v. F. BASILE, *Intelligenza artificale e diritto penale*, cit., 13 ss.

utilizzabile a fronte di condotte "seriali", come rapine, truffe agli anziani, furti in appartamento, violenze sessuali, etc.[21]; o il programma XLAW, sviluppato dalla Polizia di Napoli, e applicato in diverse regioni per prevedere furti e rapine).

Dall'altro, essi ambiscono ad individuare con maggior precisione i responsabili dei crimini commessi, *post factum*.

2.2 - *…e in sede giudiziale*

In sede *giudiziale*, il "nuovo mondo" prospetta una maggiore accuratezza delle valutazioni, attraverso indici/algoritmi predittivi che possano comunicare dati affidabili in ordine all'apprezzamento della pericolosità soggettiva (ed alla capacità di recidiva).

Si tratta, in sintesi, di algoritmi che utilizzano "*socioeconomic status, family background, neighborhood crime, employment status, and other factors to reach a supposed prediction of an individual's criminal risk, either on a scale from "low" to "high" or with specific percentages*"[22], ossia strumenti che "analizzano un numero molto elevato di dati relativi al passato e individuano delle ricorrenze (ossia dei *pattern*), caratterizzate da una base statistica molto più solida di quelle che stanno al fondo dei giudizi umani"[23].

E ciò, sia in occasione di valutazioni predittive sulla pericolosità di un soggetto condannato e ai fini della configurabilità del rischio di recidiva in sede cautelare [art. 274, lett. c), c.p.p.], o ai fini di una misura di sicurezza (art. 202 c.p.), sia in punto di commisurazione della pena (con riguardo alla capacità a delinquere di cui all'art. 133, comma 2, c.p.) o dell'applicabilità della sospensione condizionale, in sede di decisione (art. 164, comma primo, c.p.), sia in relazione alle misure alternative alla detenzione, in sede di esecuzione; senza contare il possibile utilizzo – al cospetto di analoghe valutazioni richieste – in sede di misure di prevenzione (basti pensare alla valutazione

[21] Sul punto, C. PARODI-V. SELLAROLI, *Sistema penale e intelligenza artificiale*, in *Dir. pen. cont. – Riv. trim.*, n. 6/2019, 47 ss., 56 ss.; su ulteriori strumenti "più convenzionali" di polizia predittiva v. anche G. CONTISSA-G. LASAGNI-G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Riv. trim. Diritto di Internet*, n. 4/2019, 619 ss.

[22] Così la definizione contenuta nel rapporto dell'ELECTRONIC PRIVACY INFORMATION CENTER, *Algorithms in the Criminal Justice System*, consultabile in *https://epic.org/algorithmic-transparency/crim-justice/.*

[23] M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* Risk Assessment Tools *tra Stati Uniti ed Europa*, in *www.penalecontemporaneo.it,* 28 maggio 2019, 3.

di pericolosità "per la sicurezza pubblica", presupposto della misura di prevenzione della sorveglianza speciale di pubblica sicurezza: art. 6, d.lgs. n. 159/2011).

2.3 - *La giustizia predittiva penale*

Sempre in sede *giudiziale*, ma con riferimento al momento formativo della decisione/sentenza, l'utilizzo degli algoritmi – solo a titolo di esempio – prospetta molti possibili momenti di utilizzo, promettendo:

a) di ridurre attraverso *automated decision systems* gli errori giudiziari determinati dai *bias* e dalle *fallacies* che spesso contaminano – nella fase di cognizione – la decisione giudiziale[24], nelle singole sequenze in cui si scompone – ad esempio – la valutazione della prova penale (ad es., in ordine alla valutazione della contraddittorietà di una testimonianza e dunque dell'attendibilità di un teste[25], secondo sequenze, peraltro, che superano persino le criticità alimentate dall'utilizzo di test neuroscientifici);

b) di assistere la formazione della decisione anche in sequenze intermedie, come in ordine alla prognosi favorevole di positivo sviluppo dibattimentale degli elementi di prova presentati dal PM in sede di udienza preliminare, di cui tener conto per disporre – o meno – il rinvio a giudizio;

c) di calmierare, in sede di dosimetria della pena, la *sentencing disparity*, applicando pene commisurate secondo indici di gravità calcolati oggettivamente (secondo un modello che riecheggia il *just desert model* e un aggiornamento delle *sentencing guidelines* adottate nel sistema nordamericano).

La "terra promessa" offerta dall'AI e dai *Big Data*, insomma, è quella di una "giustizia esatta", condivisa da uomini e macchine, oggi, e forse domani delegata integralmente a *justice machines*, intese come vere e proprie *trust machines*[26].

---

[24] Cfr. già C. BONA, *Sentenze imperfette*, 2010, *passim*; ed ora R. RUMIATI-C. BONA, *Dalla testimonianza alla sentenza. Il giudizio tra mente e cervello*, Bologna, 2018, in ptc. 133 ss.

[25] Cfr. ad es. A. TRAVERSI, *Intelligenza artificiale applicata alla giustizia: ci sarà un giudice* robot*?*, in *Quest. giust.*, 10 aprile 2019, 1 ss.

[26] L'ipotesi estrema della sostituzione integrale della macchina all'essere umano è quella su cui si interroga M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 2018, 872 ss.

3. L'esperienza d'Oltreoceano

Sappiamo che questo orizzonte è stato in parte anticipato – come spesso accade – dall'esperienza americana.

È noto che una sentenza della Corte Suprema del Wisconsin è intervenuta sulla decisione di un tribunale nella quale, per determinare la pena, i giudici avevano tenuto conto dei risultati elaborati dal programma COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*): uno strumento attuariale di *risk assessment* ampiamente utilizzato negli Stati Uniti per valutare il *defendant's recidivism risk*, che nel caso concreto aveva indicato l'imputato quale soggetto ad alto rischio di recidiva[27], facendo propendere per l'inflizione della pena della reclusione (a sei anni) senza *parole*, oltre a cinque anni di *extended supervision*, pena certamente elevata rispetto ai fatti marginali contestati, e per i quali si era dichiarato colpevole in sede di *guilty plea*[28].

[27] La sentenza *State v. Loomis*, 881 NW 2d 749 (Wis 2016) è stata oggetto di ampio dibattito anche sui principali media, come esempio emblematico di delegazione di scelte decisorie dell'uomo alla macchina (v. ad esempio A. LIPTAK, *Sent to Prison by a Software Program's Secret Algorithms*, in *The New York Times*, 1.5.2017).

Nella dottrina cfr., *ex multis*, I. DE MIGUEL BERIAN, *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling*, in *Law, Probability and Risk*, 2018, 45 ss.; J. LIGHTBOURNE, *Damned Lies & Criminal Sentencing Using Evidence-Based Tools*, in *Duke L. & Tech. Rev.*, 2017, 327 ss.

Nondimeno, la Corte del Wisconsin ha escluso che l'utilizzo del programma COMPAS – che aveva appunto segnalato "*a high risk of violence, high risk of recidivism, high pretrial risk*", determinando la notevole severità della sentenza di primo grado – avesse violato il diritto dell'imputato a un equo processo, sottolineando che i giudici avevano considerato legittimamente i dati forniti dal software nella determinazione della sentenza, unitamente ad altri fattori, laddove la decisione non avrebbe potuto essere basata esclusivamente o sostanzialmente sui predetti risultati. Il programma poteva quindi essere impiegato nei giudizi di determinazione della pena, con limitazioni e cautele; inoltre i punteggi di rischio non avrebbero potuto essere utilizzati come fattori determinanti nel decidere ogni qual volta il soggetto potesse essere controllato in modo effettivo e sicuro all'interno della comunità sociale (per un commento, v. *Criminal Law Sentencing Guidelines – Wisconsin Supreme Court Requires Warnings before Use of Algorithmic Risk Assessment in Sentencing – State v. Loomis*, in *Harvard Law Review*, 2017, 1530 ss.).

[28] Il programma COMPAS si basa sia su informazioni ottenute direttamente dall'imputato, in un'intervista, sia sul certificato del casellario e dei carichi pendenti, le quali vengono elaborate attraverso un modello computazionale in relazione a dati statistici di controllo, riferiti a un campione di popolazione non necessariamente corrispondente a quella dello Stato; sul piano predittivo, quindi, lo strumento prevede il rischio di ricaduta violenta, senza tuttavia offrire una spiegazione di tale rischio, individuato in rapporto al dato statistico (v. ampiamente al riguardo S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. pen.,* 2019, 1748 ss., 1750 ss.; altresì, F. BASILE, *Intelligenza artificiale e diritto penale*, cit., 19 ss., 21 ss.).

Ma programmi analoghi – come il software SAVRY – sono stati utilizzati in altre importanti decisioni[29], ed anche lo Stato del New Jersey ha sostituito le udienze per la concessione della libertà su cauzione con delle valutazioni di rischio ottenute attraverso algoritmi.

Esperienze analoghe, del resto, stanno affiorando anche nella vecchia Europa[30].

4. Le prospettive di applicazione dell'AI nel sistema penale italiano

Questo scenario, carico di potenzialità, può trovare terreno particolarmente fertile nel contesto italiano, per diverse ragioni.

In primo luogo, perché si avverte una sempre più marcata crisi della certezza del diritto, specie a fronte del crescente livello di incidenza del "precedente", e si moltiplicano gli studi dedicati alla c.d. calcolabilità del diritto, alla *foreseeability* del divieto penale ed alla *predictability* delle decisioni giudiziali, anche in ragione di una crescente *sentencing disparity*[31]. Il che alimenta l'interesse nei confronti degli strumenti di "giustizia predittiva", ossia dell'analisi di un cospicuo numero di pronunce giudiziali tramite tecnologie di AI, al fine di elaborare previsioni quanto più precise possibili in ordine al possibile esito di alcune specifiche tipologie di controversia[32].

In secondo luogo, perché il sistema penale italiano ha ormai accettato – pur con gli aggiustamenti imposti dalla Corte costituzionale – come parte integrante l'universo delle misure di prevenzione (d.lgs. n. 159 del 2011), misure cioè che hanno come componente strutturale una valutazione di pericolosità basata su dati indiziari e strutturalmente incline ad essere "integrata" da indici di carattere predittivo.

In terzo luogo perché la giurisdizione – specie nel contesto italiano – è al centro di una crisi senza precedenti, anche di legittimazione[33]: quindi è più facile predicare la

[29] V. ad esempio la decisione della Corte Suprema del *District of Columbia*, 25.3.2018, giudice Okun, citata ed analizzata ancora da S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, cit., 1754 ss.

[30] Ne dà conto M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 10 ss.

[31] V. per tutti A. CADOPPI, voce *Giurisprudenza e diritto penale*, in *Digesto/pen.*, Aggiornamento, Torino, 2016, 407 ss.; sul tema, volendo, anche V. MANES, Common law-isation *del diritto penale? Trasformazioni del* nullum crimen *e sfide prossime future*, in *Cass. pen.,* 2017, 955 ss.

[32] V. in particolare L. VIOLA, *Giustizia predittiva*, in *Enc. giur. Treccani*, Diritto on line (2018), consultabile in *http://www.treccani.it/enciclopedia/giustizia-predittiva_%28Diritto-on-line%29/.*

[33] Come del resto dimostra – dolorosamente – anche la recente *querelle* sul c.d. ergastolo ostativo, dopo che la Corte costituzionale ha ri-attribuito al magistrato di sorveglianza un potere di rivalutazione sulla pericolosità dell'ergastolano, liberandolo da automatismi legali preclusivi: il riferimento è alla sentenza della Corte costituzionale pronunciata a seguito dell'udienza pubblica del 22 ottobre 2019.

necessità di affiancare ai giudici protocolli standardizzati di giudizio in funzione tutoria, o persino invocare la necessità di sostituire (in tutto o in parte) la valutazione giudiziale con *smart machines*, meno emotive e fallibili, più razionali e prevedibili, perseguendo l'obiettivo di una giustizia terrena liberata dalle passioni e dalle imperfezioni umane.

In un contesto di crescente sfiducia nei confronti della magistratura, del resto, si può comprendere – *bon gré mal gré* – che l'opinione pubblica si senta maggiormente rassicurata da una decisone tecnica ed "automatizzata" – asseritamente "neutrale" – che da una decisione umana[34], e sia dunque favorevole ad un processo di *desimbolizzazione* della fragile umanità del diritto e del giudice e di una *resimbolizzazione* in termini scientifici[35].

D'altro canto, in questo medesimo contesto non è implausibile ipotizzare che gli stessi giudici – sempre più stretti nella morsa dell'efficienza produttiva e della crescente mediatizzazione della giustizia penale – avvertano gli algoritmi non solo come strumenti di "*reassurance juridique*" ma come provvidenziale percorso deresponsabilizzante rispetto alla mole di fascicoli affidati o a decisioni avvertite come troppo gravose o ad "alta sensibilità mediatica"[36].

5. Garanzie fondamentali e problematiche di compatibilità costituzionale

In questo scenario, dunque, i problemi che si pongono sono molteplici, e qui ci limiteremo solo ad esporre alcune preoccupazioni che l'"oracolo giuridico" garantito dagli algoritmi sollecita, muovendo nella prospettiva delle garanzie fondamentali e dei principi costituzionali.

5.1 - *Machine bias e principio di eguaglianza*

L'uso di algoritmi predittivi profila un primo problema nella prospettiva del rispetto del principio di eguaglianza (art. 3 Cost.).

[34] Si è cercato di analizzare questi aspetti in *Diritto penale* no-limits*. Garanzie e diritti fondamentali come presidio per la giurisdizione*, in *Quest. giust.*, 2019, 86 ss., 98 ss.

[35] In questi termini, ancora, A. GARAPON-J. LASSÉGUE, *Justice digitale*, cit., sui quali v. ancora E. FRONZA, *Code is law*, cit., 3.

[36] Su ulteriori "effetti di conformismo" – e di sclerotizzazione della giurisprudenza – che l'utilizzo di algoritmi può determinare, v. A. NATALE, *Introduzione. Una giustizia (im)prevedibile?*, in *Quest. giust.*, n. 4/2018, 3 ss.

In effetti, l'algoritmo – per antonomasia – è antiegualitario, perché considera alcuni fattori di rischio e non altri (età, genere, ma anche luogo di residenza, *background* socioeconomico, abitudini di vita, tendenze sessuali o "morali", data di commissione del primo precedente, etc.), e su queste basi non solo suggerisce l'allocazione di maggiori risorse di polizia in alcuni contesti urbani piuttosto che in altri, ma pone una presunzione di maggior pericolosità in relazione ad alcuni soggetti e non ad altri.

Più in generale, i risultati a cui conduce possono essere condizionati da *machine bias* non diversi da quelli che contaminano le decisioni umane[37]; e del resto – in linea con una critica diffusa proprio nel contesto di origine di queste sperimentazioni[38] – si è subito notato che "la dimensione normativa basata sulla calcolabilità del rischio e della pericolosità soggettiva apre nuovi canali di discriminazione e di incarcerazione o di ostacolo alla scarcerazione per le fasce socialmente più deboli"[39].

A questo riguardo, si è rilevato che "[…] alcuni studi hanno indicato come tali operazioni possano soffrire degli stessi pregiudizi che talvolta inficiano le decisioni umane; pregiudizi veicolati all'interno del ragionamento automatizzato dagli algoritmi utilizzati dai programmatori per il funzionamento della macchina o dal contesto all'interno del quale il *deep learning* si muove per i processi di autoapprendimento. Le ricerche si sono concentrate sul rischio che le decisioni automatizzate relative al *risk assessment* qui in esame potessero essere prese anche sulla base di un preciso pregiudizio di natura etnico-razziale, che conduceva la macchina a considerare più pericolosi gli indagati appartenenti alla comunità afro-americana. In particolare, mettendo in relazione le valutazioni di rischio svolte dall'AI e la condotta adottata dai soggetti negli anni successivi, si è evidenziato come indagati o condannati afro-americani avessero ricevuto l'indicazione di un tasso di pericolosità sociale quasi doppia rispetto a quelli

[37] Cfr. P.J. BRANTINGHAM, *The Logic of Data Bias and its Impact on Place-Based Predictive Policing*, in *Ohio St. J. Crim. L.*, 2018, 473 ss.; C. BURCHARD, *Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft*, Normative Orders Working Paper, 2/2019, 22 ss. (accessibile al link *http://publikationen.ub.uni-frankfurt.de/files/50960/Christoph_Burchard_Kuenstliche_Intelligenz_als_Ende_des_Strafrechts.pdf*).

[38] Per restare al caso del software COMPAS, sopra citato, un report dell'organizzazione *ProPublica* ha sostenuto che esso sarebbe "*biased against black*", posto che lo stesso valorizza alcuni fattori dinamici strettamente correlati alla razza: v. J. ANGWIN-J. LARSON-S. MATTU-L. KIRCHNER, *Machine Bias*, in *www.propublica.org*, 23 maggio 2016; al riguardo anche M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 5. *Amplius*, sul tema delle discriminazioni razziali nell'utilizzo di algoritmi nella giustizia penale, A.Z. HUQ, *Racial Equality in Algorithmic Criminal Justice*, in *Duke L.J.*, 2019, 1043 ss.

[39] F. SGUBBI, *Il diritto penale totale*, Bologna, 2019, 424.

caucasici; indicazione poi smentita dal comportamento concretamente tenuto negli anni successivi"[40].

5.2 - *Trasparenza e riserva di legge*

L'ingresso degli algoritmi nella struttura della decisione giudiziale – e più in generale, nella definizione della stessa area di "penalità materiale" – profila inoltre un problema di rispetto del principio democratico e di trasparenza con ricadute sensibili anche sulla tenuta del principio di riserva di legge (art. 25/2 Cost.), visto che l'algoritmo si affianca ed integra le legge, secondo il principio "*code is law*"[41].

Problemi tanto più cospicui posto che – si è autorevolmente evidenziato – "[l]'algoritmo tende a sostituire la legge. Al punto che il primato delle norme incriminatrici disposte dalla legge viene sostituito dalle norme che regolano l'applicazione del *software*: e ciò accade sia nel giudizio di fatto attinente alla individuazione di innocenza e colpevolezza dell'imputato, sia nel giudizio di diritto circa la definizione del confine fra lecito e illecito"[42].

5.3 - *Valutazioni statistiche, "diritto penale del fatto" e principio di personalità della responsabilità penale*

Sempre sul piano sostanziale, ma con risvolti anche processuali, l'uso degli algoritmi profila una modifica della cornice culturale di riferimento, rispetto ai canoni tradizionali della materialità e dell'offensività del reato, e della personalità della responsabilità penale: infatti, proiettando la valutazione del singolo caso sullo sfondo di generalizzazioni statistiche in funzione predittiva, allontana la valutazione dal *fatto*, e mette in primo piano l'autore, secondo processi di standardizzazione che prospettano una riedizione postmoderna del "tipo criminologico di autore" (*Tätertyp*).

In questa prospettiva, si è subito ammonito che "occorre evitare il rischio che attraverso questi strumenti si apra la strada a una forma inaccettabile di determinismo penale, per cui dal diritto penale del fatto – sancito dall'art. 25, comma 2, Cost. – si

[40] C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Dir. pubbl. comparato ed europeo*, 2019, 101 ss.

[41] È il titolo del saggio di L. LESSING, *Code is law. On liberty in cyberspace*, in *Harward magazine*, 2000 (mutuiamo la citazione dal citato, recente saggio di F. SGUBBI).

[42] Così ancora F. SGUBBI, *Il diritto penale totale*, cit., 41.

passi a un inaccettabile diritto penale del profilo d'autore, nel quale la pericolosità di un soggetto viene desunta esclusivamente dagli schemi comportamentali e dalle decisioni assunte in una determinata comunità del passato", che del resto "sarebbe contrario al principio di individualizzazione del trattamento sanzionatorio, desumibile dall'art. 27, commi 1 e 3, Cost., nonché, del canone di individualizzazione del trattamento cautelare, ricavabile dagli artt. 13 e 27, comma 2, Cost."[43].

5.4 - *Giusto processo e tutela del diritto di difesa*

Parallelamente, non meno significativi e rilevanti i rapporti di tensione che si avvertono nella prospettiva del "giusto processo" e del rispetto del diritto di difesa, così come del diritto ad un ricorso effettivo[44], specie nella prospettiva tradizionale che – come accennato in apertura – affida i valori in gioco ad una decisione umana, accessibile e controllabile, e maturata "nel crepuscolo del dubbio"[45].

Sotto un primo profilo, in parte speculare al precedente, si pone il problema – già emerso al cospetto del citato caso *Loomis* – della *controllabilità* dell'algoritmo, dove la possibilità di falsificare il dato elaborato da un algoritmo sta e cade – ad esempio – con la possibilità di accedere al codice sorgente che lo governa; aspetti in ordine ai quali vi è il rischio che le esigenze di tutela del segreto commerciale del programma informatico[46] finiscano col creare un "buco nero giuridico" – un *legal black hole* – e che lo strumento digitale predittivo diventi copertura di una "decisione al buio"[47].

In altri termini, il rischio è che la pronuncia guidata dal *software* predittivo sottenda una sorta di "*black box decision*", che potrebbe essere paragonata – sotto certi aspetti – ad una decisione priva di motivazione o con motivazione meramente apparente

[43] M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 21.

[44] In questa prospettiva, v. ora G. CONTISSA-G. LASAGNI-G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, cit., 620.

[45] Da questa angolatura, con un grandangolo aperto a diversi ed ulteriori problemi, v. ora le riflessioni di M. CAIANIELLO, *Criminal Process faced with the Challenges of Scientific and Technological Development*, in *European Journal of Crime, Criminal Law and Criminal Justice*, 27 (2019), 265 ss.

[46] Sul prorompere sulla scena della giustizia penale del tema dell'innovazione e delle politiche in materia di tutela del segreto commerciale, cfr. N. RAM, *Innovating Criminal Justice*, in *Nw. U. L. Rev.*, 2018, 659 ss.

[47] Si veda, nella dottrina statunitense, A.G. FERGUSON, *Illuminating Black Data Policing*, in *Ohio St. J. Crim. L.*, 2018, 503 ss.; S.B. STARR, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, in *Stan.L. Rev.*, 2014, 803 ss.

(in contrasto con l'art. 24 Cost. e con l'art. 111/6 Cost.)[48], o – peggio – ad una decisione presa nel segreto, in un contesto di asimmetria informativa (*knowledge impairment*) difficilmente compatibile con l'idea stessa di *fair trial* e con il rispetto del diritto di difesa, posto che il corollario della *equality of arms* – secondo la giurisprudenza della Corte EDU – trova la sua fondamentale essenza proprio nella facoltà di contestare e criticare le prove contrarie[49], e il suo rispetto implica che all'imputato sia stata data "*the opportunity of challenging the authenticity of the evidence and of opposing its use*"[50].

E quanto oggi sia avvertito il problema della falsificabilità del sapere tecnico introdotto nel processo penale è testimoniato – *mutatis mutandis* – dal cammino evolutivo della prova scientifica e dal definitivo abbandono della concezione veritativa della perizia: un percorso anche di recente confermato, ed ulteriormente chiarito, dalla giurisprudenza di legittimità nella sua composizione più autorevole, evidenziando "il ruolo decisivo, che, nell'ambito della dialettica processuale, assume il contraddittorio orale attraverso il quale si verifica, nel dibattimento, l'attendibilità del perito, l'affidabilità del metodo scientifico utilizzato, e la sua corretta applicazione alla concreta fattispecie processuale […], operazioni tutte che consentono anche di distinguere le irrilevanti o false opinioni del perito (cd. *junk science*) dai pareri motivati sulla base di leggi e metodiche scientificamente sperimentate ed accreditate dalla comunità scientifica"[51].

Sotto un secondo ma strettamente connesso profilo, il problema è quello dell'*affidabilità* dell'algoritmo, ossia della possibilità di adottare/accettare generalizzazioni su basi statistiche per decidere il singolo caso: l'uso degli algoritmi può infatti ingene-

[48] Ma nella prospettiva convenzionale della lesione di un diritto ad un ricorso effettivo (art. 13 CEDU e art. 47/1 CDFUE) v. i documentati rilievi critici di G. CONTISSA-G. LASAGNI-G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, cit., 620.

[49] V. ad es. il *leading case* Corte EDU, 28 agosto 1991, *Brandstetter c. Austria*, ove la Corte ha affermato che è necessario che ciascuna parte abbia effettiva conoscenza delle allegazioni e delle argomentazioni della controparte e che fruisca della concreta possibilità di contestarle e falsificarle, precisando che "*An indirect and purely hypotetical possibility for an accused to comment on prosecution argument*" non soddisfa il parametro convenzionale (§ 68); muovendo da questa decisione, per analizzare i problemi di opacità degli algoritmi e le connesse violazioni del diritto di difesa, v. ancora S. QUATTROCOLO, *Equità del processo penale e* automated evidence *alla luce della giurisprudenza della Corte europea dei diritti dell'uomo*, in *Rev. Italo-Española Der. Proc.*, vol. 2/2019, in ptc. 11 ss., 13 ss.

[50] Così la sentenza della Corte EDU, Grande Camera, 10 marzo 2009, *Bykov c. Russia*, § 90.

[51] Così, sulla traccia delle (menzionate) sentenze *Cozzini* e *Cantore* – oltre che del *leading case Franzese* – la recente SS.UU., 28 gennaio-2 aprile 2019, Pavan, sulla quale v. il commento – non privo di accenti critici – di C. BONZANO, *Le Sezioni Unite Pavan e la morte di un dogma: il contraddittorio per la prova spazza via la neutralità della perizia*, in *DPP*, 2019, 822 ss.; al riguardo, v. anche C. CONTI, *Scienza controversa e processo penale: la Cassazione e il "discorso sul metodo"*, in *DPP*, 2019, 848 ss., 860 ss.

rare effetti di sovrastima e/o rischi di "falsi positivi", e la loro scientificità e/o l'accuratezza di un "enigmatico *database*" o di un dato generato da un determinato processo computazionale dovrebbe essere fatta oggetto di valutazione peritale – non diversamente da ogni acquisizione scientifica che entri nel processo penale[52] –, e comunque essere discussa in contraddittorio nella sua fondatezza empirica, a pena di una evidente violazione – anche in questo caso – del diritto di difesa[53].

Entrambi i profili accennati evocano problematiche analoghe a quelle a cui è esposta la *prova digitale*, che qui soffrono un ulteriore aggravio dovuto al fatto che nel caso del *software* predittivo deve aversi cura di tenere concettualmente distinti ed analizzare singolarmente diversi profili: "quello della dimensione algoritmica della decisione (codificazione che trasforma un *input* in un certo *output*); quello della sua natura automatizzata; quello della validità scientifica della teoria; quello della sua tradizione in linguaggio digitale"[54].

Tutti aspetti e problemi che diventano forse insondabili "se il modello computazionale si basa su meccanismi di autoapprendimento, che portano il *software* a ricavare le regole […] non da un diagramma ad albero impostato dall'esperto, ma dall'immagazzinamento di grandi quantità di dati che gli vengono somministrati", giacché "lo stesso *designer* non è in grado di spiegare compiutamente e, quindi, di giustificare, gli *output* del modello stesso"[55].

Sotto un terzo profilo, analogamente correlato ai precedenti, emerge il problema della stessa *fruibilità* dell'algoritmo nel contesto del processo penale, e dei suoi peculiari criteri di giudizio, al di là degli ostacoli – forse non insuperabili – frapposti

[52] Ma sul punto, con ulteriori distinzioni, v. ancora S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?,* cit., 1752, a margine del citato caso *Loomis*.

[53] È ciò che sembra richiedere, ancora, la Corte EDU, Grande Camera, 10 marzo 2009, *Bykov c. Russia*, cit., § 90, quando sottolinea che "*[i]n addition the quality of the evidence must be taken in consideration, including whether the circumstances in which it was obtained cast doubt on its reliability or accuracy*"; e ciò, pur valutando la violazione dell'art. 6 § 1 CEDU alla luce dell'equità del procedimento nel suo complesso.

[54] Così, puntualmente, S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?,* cit., 1758; in ordine al problema della potenziale impossibilità di contestare l'accuratezza di una prova generata automaticamente, in assenza di una sufficiente *discovery* del modello computazionale utilizzato, v. anche U. PAGALLO-S. QUATTROCOLO, *The Impact of AI on Criminal Law and its two fold procedures*, in W. BARFIELD, U. PAGALLO, *Research Handbook on the Law of Artificial Intelligence*, 2018, Edgar Elgar, Celtenham, 395 ss.; S. QUATTROCOLO, *Equità del processo penale e* automated evidence *alla luce della giurisprudenza della Corte europea dei diritti dell'uomo*, cit., 17 ss.

[55] Ancora S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?,* cit., 1759, sulla traccia di J. BURREL, *How machines think: Understanding opacity in machine-learning algorithms*, in *Big Data and Society*, 2016, vol. 1, 1 ss.

dal divieto di perizia criminologica stabilito dall'art. 220 c.p.p. e dall'uso degli algoritmi predittivi in chiave di valutazione della pericolosità[56].

Ci si riferisce, in particolare, alle difficoltà – sul piano cognitivo e del *facts-finding* – di conciliare il criterio probatorio dell'"oltre ogni ragionevole dubbio" – la clausola BARD – con l'utilizzo di un *software*, in seno al quale *solo apparentemente* l'algoritmo fornisce maggior certezza[57]. Basti pensare al rischio di affidare l'accertamento causale – ben oltre l'apporto che si vorrebbe trarre dall'epidemiologia – a correlazioni basate su dati statistici recepite e poste a base, appunto, degli algoritmi: dove giustamente si è evidenziato che "*even with masses of data, there is no automathic tecnhique for turning correlation into causation*"[58].

5.5 - *Analisi algoritmica dei dati e diritto al silenzio dell'imputato*

C'è poi un ulteriore profilo del diritto di difesa – forse meno indagato[59], che sembra meritevole di considerazione: ci si riferisce alle frizioni che l'utilizzo di alcuni dati soggettivi – desunti attraverso una ricerca guidata da algoritmi – può determinare rispetto alla garanzia del *nemo tenetur se detegere*, ossia al diritto dell'imputato di difendersi tacendo: garanzia secolare che, come si sa, è oggetto di un rinnovato vigore nella prospettiva costituzionale e sovranazionale[60].

---

[56] Al riguardo, per ulteriori approfondimenti, v. ancora S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?,* cit., 1761 ss.

[57] Sull'approccio fideistico nei confronti del risultato fornito dagli strumenti di AI, spesso, da taluno definito *data fundamentalism*, v. in particolare K. CRAWFORD, *The hidden biases in Big Data*, in *Harward Business Review Blog Network*, 1 aprile 2013, in *https://hbr.org/2013/04/the-hidden-biases-in-big-data.*

[58] D.J. SPIEGELHALTER, *The Future lies in Uncertainty*, in *Science*, 2014, vol. 435, 264.

[59] Si veda sul punto, con riferimento all'ordinamento statunitense, C. DESKUS, *Fifth Amendment Limitations on Criminal Algorithmic Decision-Making*, in *N.Y.U. J. Legis. & Pub. Pol'y*, 2018, 237 ss.

[60] Ossia il «diritto al silenzio», «corollario essenziale dell'inviolabilità del diritto di difesa» (ordinanze n. 202 del 2004, n. 485 e n. 291 del 2002), che «garantisce all'imputato la possibilità di rifiutare di sottoporsi all'esame testimoniale e, più in generale, di avvalersi della facoltà di non rispondere alle domande del giudice o dell'autorità competente per le indagini»: diritto appunto ritenuto dalla Corte «implicito» nell'art. 24, comma secondo, Cost., così come nell'art. 6 CEDU e negli artt. 47 e 48 CDFUE (ed esplicitato solo nell'art. 14 del *Patto internazionale relativo ai diritti civili e politici*), ed in ogni caso oggetto di un autentico «concorso di rimedi» che hanno indotto la Corte costituzionale – dopo aver chiarito che tale diritto appartiene «al novero dei diritti inalienabili della persona umana […], che caratterizzano l'identità costituzionale italiana» – a disporre un rinvio pregiudiziale alla Corte di giustizia, con la recente ordinanza n. 117 del 2019, per sollecitare alcuni chiarimenti interpretativi, domandandole – in sostanza – se tale diritto si applichi, oltre che nei procedimenti penali, anche nelle audizioni personali disposte dalla CONSOB nell'ambito della propria attività di vigilanza, che può preludere all'instaurazione di procedimenti sanzionatori di natura "punitiva" nei confronti di chi sia individuato

Basti pensare – spigolando tra le voci considerate da alcuni *software* come COMPAS – all'uso di dati riferibili ad abitudini di vita, "instabilità residenziale", tendenze sessuali, gusti commerciali, "modo di utilizzo del tempo libero", ovvero a fatti pregressi privi di rilievo penale (come il "pensiero pro-criminale"), etc., tutti dati eventualmente acquisibili mediante ricerche informatiche e – nel caso – fruibili per la valutazione prognostica sulla pericolosità, o a fini analoghi, con una nuova forma di indagine retrospettiva e grandangolare che – ben oltre il perimetro chiuso e formalizzato dei precedenti penali, e superando persino il "vecchio arnese" dei pregiudizi di polizia – consente una nuova "profilazione personologica" straordinariamente invasiva, e condotta – per così dire – "*invito domino*", a tutto scapito del diritto al silenzio del soggetto[61].

E si può ulteriormente apprezzare il tono costituzionale del problema se lo si misura anche – ed ancora – nella prospettiva dell'art. 111, comma 4, Cost. (secondo il quale – come noto – "il processo penale è regolato dal principio del contraddittorio nella formazione della prova"), e della norma che ne dà attuazione, ossia l'art. 526 c.p.p., in base al quale "il giudice non può utilizzare ai fini della deliberazione prove diverse da quelle legittimamente acquisite nel dibattimento" (salvo appunto consenso dell'imputato: art. 111, comma 5, Cost.).

Senza contare problematiche, più generali, di tutela – non solo della riservatezza – in ordine ai dati personali utilizzati[62], elaborati e profilati nel *software* per alimentare il processo computazionale e l'esito predittivo (problemi oggetto di particolare attenzione in sede sovranazionale)[63]; e in ordine, più in generale, al controllo sui *Big Data* utilizzabili, in mano a pochi nuovi "Leviatani"[64].

---

come autore di un illecito (non senza evidenziare che la posizione della giurisprudenza della Corte di giustizia «non appare a questa Corte compiutamente in linea con la […] giurisprudenza della Corte europea dei diritti dell'uomo, che pare invece riconoscere una estensione ben maggiore al diritto al silenzio dell'incolpato, anche nell'ambito di procedimenti amministrativi funzionali all'irrogazione di sanzioni di natura "punitiva"»).

Per una ulteriore valorizzazione del principio, in fase esecutiva, v. ora anche la sentenza n. 253 del 2019.

[61] Per analoghe criticità poste dall'utilizzo di alcuni test neuroscientifici nel processo penale, per verificare l'attendibilità di testimoni, vittime e imputati, cfr. ancora M. CAIANIELLO, *Criminal Process faced with the Challenges of Scientific and Technological Development,* cit., 280 ss.

[62] Al riguardo, per un inquadramento generale dei problemi, v. ad es. G. FINOCCHIARO, *Intelligenza artificiale e protezione dei dati personali*, in *Giur.it.*, 2019, 1670 ss., cui si rinvia anche per i riferimenti alla recente normativa UE, anche con riguardo al "diritto di spiegazione" in ordine a "decisioni autonomizzate", ed al problema della "qualità dei dati".

[63] Sui profili di contrasto con l'art. 8 CEDU v. ancora S. QUATTROCOLO, *Equità del processo penale e* automated evidence *alla luce della giurisprudenza della Corte europea dei diritti dell'uomo*, cit., 5 ss.

[64] Cfr. ancora C. CASONATO, *Intelligenza artificiale e diritto costituzionale*, cit., 101 ss.

6. Riflessioni conclusive: l'algoritmo come supporto alla decisione giudiziale

Questi problemi, certo, non devono apparire una chiusura radicale all'utilizzo dell'AI e degli algoritmi in materia penale, perché sarebbe irragionevole – oltre che antistorico – rinunciare alle componenti positive che lo sviluppo tecnologico offre al sistema penale: non c'è dubbio, infatti, che "[…] l'ausilio della nuova tecnologia digitale, di *software* informatici e algoritmi predittivi sull'esito delle controversie, o […] l'apporto della robotica e della logica dell'intelligenza artificiale, possono certamente contribuire alla correzione dell'utilizzo improprio di euristiche fuorvianti"[65], ossia delle procedure semplificate (*heuristic*) e gli automatismi fuorvianti (*biases*) che spesso condizionano, in concreto, le decisioni giudiziali, e della "razionalità limitata" – "*fast and frugal"* – che le caratterizza.

Alcune rilevazioni empiriche, d'altronde, segnalano la proficuità di un impiego degli algoritmi come ausilio e riscontro rispetto alla decisione giudiziale: così, ad esempio, "[…] uno studio pubblicato nell'agosto del 2017, condotto sull'analisi retrospettiva di oltre 758.000 casi di decisioni relative al rilascio su cauzione adottate fra il 2008 e il 2013 nel distretto di New York, ha dimostrato come l'utilizzo di algoritmi avrebbe potuto permettere la riduzione dell'indice di criminalità del 24.7 % mantenendo inalterato il tasso di detenzione. Altrimenti avrebbe potuto condurre ad una riduzione del 41.9% del tasso di carcerazione senza per questo aumentare i livelli di criminalità"; ed al riguardo si è evidenziato che "[…] una conseguenza sull'uso degli algoritmi, sulla base di una impostazione basata su "*constructing unbiased decision counterfactuals*", sarebbe stata anche una riduzione delle disparità di trattamento legate all'origine razziale delle persone coinvolte"[66].

L'orizzonte, dunque, dovrebbe essere quello che – confermando il perentorio divieto di una decisione basata *unicamente* su trattamenti automatizzati[67] – ammetta una funzione *solo* tutoria dell'algoritmo per limitare la fallibilità della decisione giudiziale, nel quadro di una collaborazione uomo-macchina.

---

[65] R. RUMIATI-C. BONA, *Dalla testimonianza alla sentenza*, cit., 11.

[66] C. CASONATO, *Intelligenza artificiale e diritto costituzionale*, cit., 105.

[67] Divieto già da tempo fissato in sede europea, e ribadito dall'art. 22 del Regolamento (UE) n. 2016/679, prevedendo che "l'interessato ha il diritto a 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" (primo comma); diritto aperto, tuttavia, a deroghe e limitazioni (art. 22, comma secondo, e 23, Reg. citi.): al riguardo, v. ancora G. FINOCCHIARO, *Intelligenza artificiale e protezione dei dati personali*, cit,, 1674 s.; ma sul punto, anche per taluni accenti critici, v. ancora M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 16 ss.

Così, con riferimento all'utilizzo di algoritmi predittivi in ordine al giudizio di pericolosità o recidiva, si è ipotizzato di assegnare all'*output* prodotto dall'AI valenza di *mero indizio*, bisognoso di essere sempre corroborato con altri elementi di prova[68]; ma può essere anche plausibile ipotizzare – simmetricamente – l'apporto argomentativo offerto dai *software* predittivi come strumento di *double check*, a valle dunque della valutazione del giudice.

In questa precipua prospettiva, in altri termini, si potrebbe prevedere un percorso dove l'algoritmo possa essere impiegato come strumento di verifica della scelta operata dal giudice, e di confronto con l'esito offerto da una valutazione informatizzata e verificata attraverso i dati; e ipotizzare un obbligo di motivazione rafforzata – sulla traccia di quello via via emerso al cospetto della prova scientifica, a partire dalla sentenza Cozzini[69], nella giurisprudenza della Cassazione[70] – nel caso in cui, ad esempio, il giudice che ha ritenuto pericoloso/a rischio recidiva un soggetto mediante una valutazione che l'algoritmo smentisce, sia costretto a confutare con argomentazioni di "efficacia scardinante" o di "forza persuasiva superiore"[71] questa *second opinion* "artificiale" – valevole almeno a far sorgere un "ragionevole dubbio" – ove voglia confermare il giudizio prognostico negativo[72].

---

[68] M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 17 s., rilevando – anche alla luce di una peculiare interpretazione della normativa UE di riferimento – che "l'interessato ha diritto a che sul suo *status* si pronunci un giudice in carne ed ossa, che dovrà tener conto *anche* di elementi di prova ulteriori rispetto all'*output* del *risk assessment tool*", traendone la conseguenza che "[n]on vi è spazio dunque in Europa per uno scenario analogo a quello della California e del Kentucky".

Anche secondo la Suprema Corte del Wisconsin, nel caso *Loomis*, del resto, il programma COMPAS – pur riconosciuto utilizzabile – "*should always constitute merely one tool available to a Court, that need to be confirmed by "additional sound informations*"" (*Loomis v. Wisconsin*, July 13, 2016): v. sul punto, da ultimo, M. CAIANIELLO, *Criminal Process faced with the Challenges of Scientific and Technological Development,* cit., 283 ss.

[69] Sulla quale, tra gli altri, v. P. TONINI, *La Cassazione accoglie i criteri* Daubert *sulla prova scientifica. Riflessi sulla verifica delle massime di esperienza*, in *DPP*, 2011, 1345 ss.; più di recente, R. BLAIOTTA-G. CARLIZZI, *Libero convincimento, ragionevole dubbio e prova scientifica*, in AA.VV., *Prova scientifica e processo penale*, a cura di G. Canzio-L. Luparia, cit., 367 ss.; G. CARLIZZI, Iudex peritus peritorum, in *Dir. pen. cont. – Riv. trim.*, n. 2/2017, 28 ss.

[70] Per una istruttiva disamina, cfr. C. CONTI, *Scienza controversa e processo penale: la Cassazione e il "discorso sul metodo"*, cit., 853 ss.; più in generale, sull'evoluzione dell'approccio al problema, v. il suggestivo affresco di G. CANZIO, *La motivazione della sentenza e la prova scientifica:* "reasoning by probabilities", in AA.VV., *Prova scientifica e processo penale*, cit., 3 ss.

[71] Torna sul tema della motivazione rafforzata, di recente, M. CECCHI, *La "motivazione rafforzata" del provvedimento ovvero la "forza persuasiva superiore"*, in *DPP*, 2019, 1123 ss.

[72] Dovendo dunque assolvere l'obbligo – per parafrasare la nota sentenza della SS.UU., *Mannino* – di delineare le linee portanti del proprio, alternativo, ragionamento, e di confutare specificamente i più rilevanti argomenti a sostegno della diversa valutazione predittiva operata dal *software*.

Nel caso inverso, ove una valutazione predittiva positiva del giudice sia smentita dall'algoritmo, il *focus* della analisi giudiziale – prima di attribuirle credito – dovrebbe concentrarsi sulla "scientificità" dell'algoritmo, sino ad affermare "che esso risponde ai canoni di verificabilità della prova scientifica, avuto riguardo ai princìpi della controllabilità, della falsificabilità e della verificabilità della teoria posta a fondamento della prova"[73], ed eventualmente confrontare l'esito predittivo con quello offerto da altro e diverso *software* (anche se, ove volesse confermare la prognosi di non pericolosità, potrebbe essere non implausibile considerare *non* strettamente necessaria una motivazione rafforzata, in base all'*in dubio pro libertate*).

L'aggravio dell'attività giudiziale sarebbe, del resto, controbilanciato dal guadagno in punto di affidabilità della stessa, e l'intervento umano corredato da questo peculiare onere motivazionale – al cospetto di una decisione *algorithm based* – potrebbe ridurre la frizione con le garanzie dell'equo processo e con il diritto ad un ricorso effettivo[74].

In ogni caso, la valutazione del giudice resterebbe doverosamente centrale, specie quando ha ad oggetto la personalità dell'imputato, come del resto indicato dalla Corte costituzionale quando nell'indagare la *ratio* del divieto di perizia criminologica (art. 220, comma secondo, c.p.p.) vi ha scorto (anche) la preoccupazione "che lo studio della personalità dell'imputato possa venire compiuto *solo da chi abbia presente anche il carattere afflittivo e intimidatorio della pena*"[75].

In definitiva, anche nella prospettiva dei rapporti tra AI e diritto penale va considerata con sorvegliata attenzione la posizione di chi sostiene che "alcune qualità umane non potranno mai essere sostituite da componenti artificiali: si tratta, a seconda

[73] Cfr. Cass., sez. I, 23 novembre 2018, B., est. S. Aprile, sul caso di Yara Gambirasio: al riguardo, v. ancora C. CONTI, *Scienza controversa e processo penale*, cit., 853 ss.

[74] Ma sul punto v. però i dubbi di G. CONTISSA-G. LASAGNI-G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo* cit., 630, secondo i quali "[…] la mera previsione di un successivo intervento umano non sembra sufficiente a garantire un rimedio effettivo, specialmente alla luce dei parametri di "effettività" utilizzati dalle due Corti europee".

Tali AA. sarebbero, viceversa, favorevoli a un *double check* sempre mediante uno strumento di AI (secondo un metodo già invalso in altri settori *safety critical*, come l'aviazione), avendo cura di precisare che "[n]on sarebbe sufficiente che entrambi i sistemi siano certificati e validati per l'utilizzo nel processo penale; per garantire un rimedio effettivo, questi dovrebbero anche essere progettati e sviluppati da produttori diversi", garantendo una diversità di approccio – secondo il c.d. *redundancy approach* – "che potrebbe essere ottenuta adottando approcci alternativi nello sviluppo degli algoritmi, impiegando diversi team di programmatori e selezionando diversi componenti *hardware* e *software*".

[75] Così la sentenza n. 124 del 1970 (corsivi nostri), giustamente richiamata, ancora, da M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., 22.

delle impostazioni, dell'immaginazione; della capacità di dare vita a processi creativi; della coscienza, intesa secondo la teoria dell'informazione integrata; di creatività, emozioni e ispirazione, frutto dell'azione degli ormoni. E anche il beneficio del dubbio, con il correlato senso di curiosità, e la sana consapevolezza di sapere di non sapere sono caratteristiche che contraddistinguono l'umano e la sua ricerca di senso, le quali mal si attagliano ai ragionamenti dell'AI"[76].

Si avvertono con inquietudine – conclusivamente – lo *shock* epistemologico ed antropologico ed i rischi che sarebbe chiamato ad affrontare un sistema di giustizia penale che progetti di sostituire con i prodigi dell'AI la componente umana dello *ius dicere*, sino a prefigurare una "rivoluzione numerica" dove si possa abbandonare la "liturgia della parola" e "decidere senza giudicare"[77], e uno scenario dove la tecnologia – liberata dal fattore umano – diviene tecnocrazia: rischi di fronte ai quali appare preferibile conservare un modello che – senza rinunciare ai correttivi offerti dal progresso tecnologico – continui a perseguire l'obiettivo di una "giustizia *giusta*", rispetto all'utopia *legaltech* di una "giustizia *esatta*".

[76] Così ancora C. CASONATO, *Intelligenza artificiale*, cit., 124, che su queste basi giunge ad affermare il *diritto di conoscere la natura umana o artificiale dell'interlocutore* e, soprattutto, il *diritto ad una decisione umana*; in una prospettiva non distante, v. ancora le preoccupazioni di M. CAIANIELLO, *Criminal Process faced with the Challenges of Scientific and Technological Development,* cit., 288 ss.

[77] Al riguardo, v. anche le perplessità di C. CASONATO, *Intelligenza artificiale*, cit., 118 ss.