

## EDITORIALE

---

**ANNA MARIA MAUGERI**

### **L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra *evidence based practices* e tutela dei diritti fondamentali.**

Il lavoro analizza in chiave problematica le prospettive legate all'utilizzo di algoritmi predittivi per valutare la pericolosità sociale, nella consapevolezza dei vantaggi che possono derivare da un sistema di accertamento evidence - based, da una parte, e dei rischi per la tutela dei diritti fondamentali, dall'altra; l'algoritmo può rappresentare una perizia, sottoposta al vaglio del giudice adottando i criteri "Daubert"?

*The use of predictive algorithms to ascertain social danger: a challenge between evidence based practices and protection of fundamental rights.*

*The article analyses in a problematic key the perspectives linked to the use of predictive algorithms to assess social dangerousness (as risk assessment tools), in the awareness of the advantages that can derive from an evidence-based assessment system, on the one hand, and of the risks for the protection of fundamental rights, on the other; can the algorithm represent an expert testimony, submitted to the judge by adopting the "Daubert" criteria?*

**SOMMARIO:** 1. Premessa. - 2. I vantaggi dell'uso di algoritmi predittivi contro l'incertezza del giudizio di pericolosità sociale. - 3. I rischi connessi all'uso di algoritmi predittivi: discriminazioni e valutazioni generalizzanti. - 4. La mancanza di trasparenza. - 5. L'adozione del modello "Daubert" per verificare la scientificità della teoria *psico-criminologica* e del software. - 6. Il diritto di "accesso" delle parti interessate all'algoritmo nella decisione del Consiglio di Stato, n. 2270/2019. - 7. Il giudizio individualizzato e il controllo del giudice sull'output. - 8. Le prospettive relative all'uso dell'algoritmo in *bonam partem*.

1. *Premessa.* Nei più recenti studi circa l'utilizzo dell'intelligenza artificiale nell'ambito del sistema di giustizia penale emerge l'uso dei c.d. algoritmi predittivi a servizio del giudice per la capacità di calcolare il rischio di reiterazione del reato sia in sede di commisurazione della pena sia ai fini dell'applicazione delle misure cautelari o alternative.

Algoritmi sono già ampiamente utilizzati negli Stati Uniti, basti pensare all'uso del sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)<sup>1</sup> che ai fini del risk assessment in sede di commisurazione

---

<sup>1</sup> Cfr. QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, in *BioLaw journal = Rivista di BioDiritto*, 1, 2019, 135 ss. Tra i software privati più evoluti si annovera COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) che processa in input variabili riconducibili a cinque diverse categorie: precedenti penali, personalità e stile di vita, attitudini personali, composizione familiare, emarginazione sociale. Nelle giurisdizioni di molti Stati, tra cui Wisconsin, Florida e Michigan, questo software viene costantemente utilizzato per supportare i giudici nel sentencing. Il sistema "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, ma in rapporto al dato

della pena utilizza algoritmi basati su numerosi dati personali<sup>2</sup> o il Public Safety Assessment (PSA), utilizzato da 28 giurisdizioni (e, tra loro, tre Stati) che assiste il giudice nella decisione circa il rischio di fuga dell'imputato prima del processo<sup>3</sup>.

Algoritmi, allora, potrebbero essere utilizzati per l'accertamento e valutazione della pericolosità sociale, basata proprio su una valutazione prognostica circa la probabilità che il soggetto torni a delinquere o (ai fini delle misure di prevenzione) delinqua *tout court*, e, quindi, sia in sede di applicazione delle misure di sicurezza (art. 202 c.p.) sia delle misure di prevenzione (art. 6 d.lgs. 159/2011)<sup>4</sup>, ma anche in occasione di valutazioni predittive sulla pericolosità di un soggetto ai fini della configurabilità del rischio di recidiva in sede caute-

---

statistico". "Si tratta di uno strumento attuariale che valuta il rischio statico, non dinamico: infatti, gli strumenti attuariali non spiegano il recidivismo, si limitano a segnalarlo, valutando i fattori di rischio attraverso statistiche ufficiali e prospettive teoriche comprensive. Sul mercato, lo strumento è commercializzato in forma di software, da Northpointe inc. che ne detiene i diritti e le licenze commerciali. Gli strumenti di risk assessment, però, non sono necessariamente dei software" (QUATTROCOLO, *op. cit.*, nota 33).

<sup>2</sup> In materia cfr. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2020, 18 ss.; QUATTROCOLO, *Equo processo*, *cit.*, in particolare 131 ss.; UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale 2020*; PARODI - SELLAROLI, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Dir. pen. cont.*, 2019, n. 6, 66 ss.; MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Dir. pen. cont. Riv. Trin.* 2019, 2, 401 ss.; DI GIOVINE, *Il judge-bote le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziali)*, in *Cass. pen.* 2020, 3, 951; D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont.* 2019, 2, 359 ss.

<sup>3</sup> [www.psapretrial.org](http://www.psapretrial.org), in cui viene spiegato quali siano i fattori utilizzati e il peso di ciascun fattore, rendendone pubblico il funzionamento. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for a European Legal Discussion*, Springer (Cham, Switzerland) 2020, (nel prosieguo *op. cit.*), 151 ss.: "The risk factors considered by PSA under this parameter are: Pending charge at the time of offense (Y/N); Prior conviction (Y/N); Prior failure to appear in the past 2 years (No/Once/2 or more times); Prior failure to appear, older than 2 years (Y/N)". L'uso di questo *software*, elaborato nel 2013 doveva essere utile per ridurre il numero di individui detenuti prima del processo, ma i risultati sono considerati scarsi, nonostante ciò "the movement for bail reform is one rare area of bi-partisan agreement, pushing for abandoning monetary bail, in favour of effective management of risk". Cfr. FAZEL, *The scientific validity of current approaches to violence and criminal risk assessment*, in De Keijser, Roberts, Ryberg (eds), *Predictive sentencing, normative and empirical perspective*, Oxford, 2019, 197, da cui emerge che 39 Stati federali hanno il loro "risk assessment tool"; in Inghilterra e in Galles, OASys è il sistema di valutazione degli autori di reato utilizzato abitualmente nel National Offender Management Service (NOMS), all'interno del quale vengono impiegati diversi strumenti di valutazione del rischio, a seconda delle esigenze criminogene dell'autore del reato valutato. Cfr. CONTISSA, LASAGNI, SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Dir. di Internet* 2019, 4, 621 s.

<sup>4</sup> Cfr. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Studi in onore di Lucio Monaco*, a cura di Bondi, Fiandaca, Fletcher, Marra, Stile, Roxin, Volk, Urbino 2020, 687.

lare [art. 274, lett. c), c.p.p.], in sede di applicazione della sospensione condizionale (art. 164, comma primo, c.p.) o delle misure alternative alla detenzione, in sede di esecuzione, nonché, al limite, in sede di commisurazione della pena con riguardo alla capacità a delinquere di cui all'art. 133, comma 2, c.p.. L'utilizzo di algoritmi viene esteso anche alla scelta del tipo di programma di riabilitazione in taluni settori come quello dei sex offenders, che già da epoca risalente è stato basato in molti ordinamenti, innanzitutto in Canada, sulla valutazione del rischio di recidiva con sistemi basati sul Risk-Need-Responsivity<sup>5</sup> (RNR, il trattamento deve essere proporzionale al rischio di commettere un nuovo reato), che utilizzano strumenti attuariali, basati su fattori di rischio validati empiricamente, tratti dalla storia personale e criminale del soggetto (STATIC 99R, STABLE 2007, ACUTE)<sup>6</sup>. Attraverso il modello RNR, gli strumenti di valutazione del rischio sono tra le applicazioni più comuni della tecnologia di Intelligenza Artificiale alla giustizia penale, secondo il 2018 *Global Meeting on the Opportunities and Risks of AI and Robotics for Law Enforcement*. Questi strumenti, già fortemente utilizzati nei "correctional and probation services" occidentali, da una parte calcolano, in base all'individuazione e alla pesatura di fattori di rischio statici (ad esempio la storia criminale) e dinamici, il rischio o la probabilità di recidiva dell'individuo, per la criminalità in generale e/o per tipi specifici di reato, come appunto la criminalità sessuale; dall'altra parte, questa valutazione del rischio viene utilizzata allo scopo di adattare il "trattamento" alla necessità di modificare i fattori di rischio dinamici presentati dall'individuo, o rispondere alle esigenze criminogene dell'individuo, in carcere o in libertà vigilata, nonché influenzare il processo decisionale in materia di libertà vigilata e monitorare l'individuo dopo il rientro nella comunità<sup>7</sup>.

L'uso di simili strumenti, perlomeno in termini di valutazione del rischio di recidiva con sistemi basati sul Risk-Need-Responsivity, si incomincia a pro-

---

<sup>5</sup> BONTA, ANDREWS, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation* 2007-06,

[www.courtinnovation.org/sites/default/files/documents/RNRModelForOffenderAssessmentAndRehabilitation.pdf](http://www.courtinnovation.org/sites/default/files/documents/RNRModelForOffenderAssessmentAndRehabilitation.pdf); sul rischio di recidiva nel settore in esame cfr. XELLA, *La valutazione del rischio di recidiva per gli autori di reati sessuali*, in *Riv. it. med. leg. e dir. in campo sanitario* 2013, 1885 ss..

<sup>6</sup> XELLA, *La valutazione del rischio di recidiva per gli autori di reati sessuali*, in *Riv. it. med. leg. e dir. in campo sanitario* 2013, 1891 ss.; RAFFAELE, *La recidiva dei reati sessuali e la castrazione chimica*, in *Trattato di diritto e bioetica*, a cura di Cagnazzo, Napoli 2017, 587 ss.

<sup>7</sup> PEREIRA, *Artificial Intelligence, Offender Rehabilitation & Restorative Justice*, febbraio 2020-02, in *The "Good" Algorithm? Artificial Intelligence: Ethics, Law, Health. International Workshop organised by the Pontificia Academia Pro Vita*, Date: 2020/02/26 - 2020/02/28, Location: New Hall of the Synod, Vatican City, [liras.kuleuven.be/2960856?limo=0](https://liras.kuleuven.be/2960856?limo=0).

spettare anche nell'ordinamento italiano; ad esempio, strumenti di valutazione del rischio di recidiva come *STATIC 99R*, *STABLE 2007* e *ACUTE* sono utilizzati in un programma condotto dal CIPM nel carcere di Milano Bollate e presso il Presidio criminologico del Comune di Milano dal 2005, per la scelta del tipo di trattamento nei confronti degli autori di reati sessualmente connotati<sup>8</sup>. O ancora in una recente sentenza del Tribunale di Milano si fa riferimento all'uso di strumenti di radice nordamericana per la valutazione del rischio di recidiva, funzionale alla valutazione della pericolosità sociale ai fini dell'applicazione di una misura di sicurezza nel settore dei reati sessuali: «Tali elementi sono condivisi dall'uomo in sede di colloqui individuali oppure durante l'indagine del rischio di recidiva, realizzato con strumenti di matrice nord americana, che indicano un livello globale “medio-alto” (circa il 25% di possibilità di commettere un nuovo reato sessuale nei 5 anni dopo la detenzione)»<sup>9</sup>.

La rilevanza della questione è testimoniata anche dall'interesse del Consiglio d'Europa per il crescente impiego di strumenti digitali, governati, in generale da algoritmi e forme più o meno sofisticate di intelligenza artificiale, nell'ambito della giurisdizione, come testimoniato dapprima dalla pubblicazione dello studio *Algorithms and Human Rights* (DGI (2017)12) nel marzo del 2018 e poi nel dicembre dello stesso anno della “Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti”. La *Carta* pretende che l'elaborazione e l'attuazione di strumenti e servizi di intelligenza artificiale siano conformi a una serie di principi, tra cui, innanzitutto, il principio del rispetto dei diritti fondamentali, rivolto in prima battuta agli operatori del settore privato, come sono solitamente gli sviluppatori di *software*; la Carta richiama così sia la Convenzione europea dei diritti dell'uomo sia la Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati di carattere personale (adottata a Strasburgo il 28 gennaio 1981 e resa esecutiva in Italia con L. 21 febbraio 1989 n. 98), che stabiliscono alcuni principi considerati rilevanti nell'attività di creazione del *software*, quale strumento, tra l'altro, di supporto al *decision-making* giurisdizionale: il principio di legalità; il diritto di accesso alla giurisdizione; il diritto al processo equo, nelle sue articolazioni essenziali del contraddittorio e della

---

<sup>8</sup> XELLA, *op. cit.*, 1887.

<sup>9</sup> Tribunale Milano, 19 aprile 2016, in *www.dejure.it*.

parità delle armi; l'indipendenza della magistratura e, in particolare, dei giudici nell'esercizio del potere decisorio<sup>10</sup>.

Nell'ambito dell'Unione Europea, tra l'altro, la Commissione ha elaborato il "Libro bianco sull'IA. Un approccio europeo all'eccellenza e alla fiducia", del 19 febbraio 2020, (COM(2020) 65) che prevede che gli sviluppatori di sistemi di IA sono già tenuti a rispettare le norme europee in tema di diritti fondamentali, protezione dei consumatori, sicurezza dei prodotti e responsabilità, ferma restando la necessità di implementare nuove regole per far fronte ai nuovi rischi connessi all'IA. La Commissione propone una differenziazione sul piano della disciplina, distinguendo tra applicazioni di IA che comportano un alto rischio e applicazioni a basso rischio, con la conseguenza che solo per le prime sarebbe imposta l'osservanza di talune norme e si sarebbe tenuti a dimostrare il possesso di determinati requisiti. Un particolare interesse assume, poi, lo studio pubblicato nel luglio 2020 su «Artificial Intelligence and Law Enforcement. Impact on fundamental Rights», commissionato dal Parlamento europeo sull'impatto sui diritti fondamentali, e in particolare il diritto alla privacy, dell'utilizzo di strumenti di IA nell'ambito delle attività di polizia e della giustizia penale, il tutto dal punto di vista dell'Unione Europea<sup>11</sup>.

La recentissima proposta di Regolamento "Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts", 21 aprile 2021, esprime la consapevolezza delle potenzialità connesse all'uso di strumenti di IA nell'ambito del "criminal law enforcement" ma, allo stesso tempo, considera tali strumenti ad alto rischio per i possibili impatti negativi sui diritti fondamentali garantiti nella Carta, soprattutto laddove utilizzati per valutazioni individuali, dai poligrafi e strumenti simili, per accertare lo stato emotivo di una persona, per rilevare "deep fakes", per valutare l'attendibilità delle prove nei procedimenti penali, e, con particolare riferimento al settore in esame (valutazione della pericolosità sociale), "*for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons, or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups, for profiling in the course of detection, investigation or*

---

<sup>10</sup> *Intelligenza artificiale e giustizia: nella cornice della carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Leg. pen.* 2018, 6.

<sup>11</sup> [www.europarl.europa.eu](http://www.europarl.europa.eu).

*prosecution of criminal offences, as well as for crime analytics regarding natural persons”* (corsivo aggiunto) (considerando n. 38)<sup>12</sup>.

2. *I vantaggi dell’uso di algoritmi predittivi contro l’incertezza del giudizio di pericolosità sociale.* Le nuove tecnologie offrono il vantaggio di poter attingere ed elaborare quantità enormi di dati da fonti quali banche-dati giurisprudenziali, legislative, raccolte di precedenti, e simili – con l’utilizzo di dispositivi molto sofisticati, che dovrebbero consentire di far emergere *relazioni, coincidenze, correlazioni*, che consentano di profilare una persona e prevederne i successivi comportamenti, anche di rilevanza penale.

Considerando che un algoritmo è una sequenza di istruzioni che deve essere seguita per trasformare un *input* in un *output*<sup>13</sup>; sono 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”<sup>14</sup>. Il *machine learning*, a sua volta, consiste in un algoritmo che è in grado di apprendere partendo dai dati e riesce a affinare le proprie conoscenze attraverso l’esperienza; grazie al *machine learning* si tenta di imparare dal passato qualcosa che possa valere anche per il futuro<sup>15</sup>. L’apprendimento è finalizzato alla predizione, alla risoluzione di casi diversi da quelli analizzati, che si potranno verificare in futuro<sup>16</sup>. Questo processo è definito *generalizzazione*: dato un insieme di informazioni iniziali, bisogna estrapolare una regola che sia idonea a prevedere e risolvere casi futuri non ancora analizzati<sup>17</sup>: il *machine learning* mira a prevedere un certo risultato. Secondo una nota definizione, “*A computer program is said to learn from experience E with respect to some class of tasks T and*

---

<sup>12</sup> Nel titolo IV, intolato *Transparency Obligations For Certain Ai Systems* si prevede che l’obbligo di informare le persone fisiche che stanno interagendo con sistemi di IA non si applica nel caso di sistemi autorizzati dalla legge per rilevare, prevenire, indagare e perseguire reati, tranne laddove tali sistemi siano disponibili al pubblico per denunciare un crimine.

<sup>13</sup> ALPAYDIN, *Introduction to Machine Learning*, The MIT Press, Cambridge, 2010, 1.

<sup>14</sup> 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](http://www.penalecontemporaneo.it).

<sup>15</sup> DOMINGOS, *L’algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, Bollati Boringhieri, cit., 86.

<sup>16</sup> Così GIUCA, *Introduzione sui sistemi di intelligenza artificiale*, cap. I, 7, inedito.

<sup>17</sup> Riportano questa definizione in linguaggio matematico BELKIN, HSU, MA, MANDAL, *Reconciling modern machine learning practice and the classical bias-variance trade off*, in *PNAS*, vol. 116, 32, 2019, 158.

*performance measure P, if its performance at tasks in T, as measured by P, improves with the experience E<sup>18</sup>.*

Tali dispositivi potrebbero agire con una metodologia oggettiva e priva di pregiudizi e il vantaggio dell'utilizzo di algoritmi predittivi potrebbe essere rappresentato da una maggiore certezza del diritto, mentre sino ad ora la valutazione della pericolosità sociale, soprattutto nella parte prognostica del giudizio (predittiva rispetto al futuro), consiste per lo più in giudizi intuitivi, affidati all'esperienza personale dei giudici e al loro buon senso, salvo l'intervento di perizie laddove consentite<sup>19</sup>. Il giudizio di pericolosità sociale ai fini dell'applicazione delle misure di sicurezza, infatti, presenta i gravi limiti di affidabilità scientifica dei giudizi criminogenetici e predittivi, al punto che, si osserva, «può fornire basi pseudo-scientifiche a quelle forme di discriminazione – sociale, politica, culturale, religiosa, razziale – così frequenti nel crepuscolo delle democrazie»; tanto è vero che non stupisce che il diritto penale di questi ultimi anni, modulandosi sul telos della sicurezza<sup>20</sup>, proiezione meta-oggettiva della pericolosità, abbia frequentemente individuato proprio nei più deboli i suoi bersagli elettivi: negli emigranti, negli emarginati, negli *homeless*, nelle prostitute di strada, in quelli che fanno graffiti, nei tossici<sup>21</sup>. Il concetto di pericolosità, infatti, si configura come un “ibrido”, connotato contemporaneamente da parametri medici e da parametri giuridici, molto ambiguo e scientificamente inconsistente<sup>22</sup>. La stessa Suprema Corte è cosciente del fatto

<sup>18</sup> MITCHELL, *Machine Learning*, McGraw Hill, 1997, 2; ampiamente cfr. FABBRI, *Cos'è l'intelligenza artificiale e quali sono le applicazioni attuali e future*, in [www.zerounoweb.it](http://www.zerounoweb.it).

<sup>19</sup> Cfr. CONTISSA, LASAGNI, SARTOR, *op. cit.*, 631; BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. It. di dir. e proc. pen.*, 2019, 1926 il quale evidenzia lo spostarsi della fiducia dalle persone alla tecnologia. Sulla problematicità del giudizio prognostico di pericolosità cfr. gli autori citati in nota 330.

<sup>20</sup> CAVALIERE, *Può la “sicurezza” costituire un bene giuridico o una funzione del diritto penale?*, in *In dubio pro libertate. Festschrift für Klaus Volk zum 65. Geburtstag*, a cura di Hassemer, Kempf, Moccia, München, 2009, 111.

<sup>21</sup> BALBI, *Infermità di mente e pericolosità sociale tra opg e rems*, in *Dir. pen. cont.* 20-7-2015.

<sup>22</sup> SALVATI, *La pericolosità sociale nell'ordinamento giuridico italiano*, in *Amministrazione in cammino*, 11 maggio 2011; MASTRONARDI, *Manuale per operatori criminologici e psicopatologi forensi*, Milano, 1996, 388; sulla problematicità della prognosi di pericolosità sociale, tra gli altri, BERTOLINO, *Il “crimine” della pericolosità sociale: riflessioni da una riforma in corso*, in *Riv. it. med. leg.* 2016, 1371 ss.; FIANDACA, *L'imputabilità nell'interazione tra epistemologia scientifica ed epistemologia giudiziaria*, in *Leg. pen.* 2006, 263; MANNA, *Imputabilità tra prevenzione generale e principio di colpevolezza*, in *Leg. pen.* 2006, 241; Id., *Imputabilità, pericolosità e misure di sicurezza: verso quale riforma?*, in *Riv. it. dir. proc. pen.*, 1994, 1426 ss.; COLLICA, *La crisi del concetto di autore non imputabile “pericoloso” del reato*, in *Dir. pen. cont.*, 2012, 274 ss.; PADOVANI, *La pericolosità sociale sotto il profilo giuridico*, in *Trattato di criminologia*, a cura di Ferracuti, vol. XIII, 318 ss.; PETRINI, *La prevenzione inutile*, Napoli, 1996, 294 s.; MARTINI, *Essere pericolosi. Giudizi soggettivi e misure personali*, Torino, 2017, 155;

che «la valutazione della pericolosità sociale (art. 203 c.p., che richiama al comma 2 l'ineludibile declinazione operata dal legislatore all'art. 133 c.p.) è, come ogni giudizio prognostico, fondato sull'apprezzamento della ricorrenza di un "pericolo", che è per sua natura un giudizio rivolto al futuro, *il che esclude una sua possibile declinazione in termini di certezza storica* (attributo con cui si possono, convenzionalmente e processualmente, qualificare solo condotte passate), *con un ineliminabile margine di fallibilità*» (corsivo aggiunto)<sup>23</sup>.

Anche nel settore delle misure di prevenzione, essendo in discussione una qualità della persona e non fatti concreti, il giudizio di pericolosità sociale rimane un giudizio di probabilità<sup>24</sup> dal quale è esclusa ogni certezza; si tratta di «enunciazioni di probabilità sul futuro comportamento legale dei soggetti»<sup>25</sup>, che risolvendosi in un giudizio a struttura probabilistica non possono svolgersi che per approssimazione<sup>26</sup>, a parte la precarietà dei giudizi fondati su prognosi<sup>27</sup> (si tratta di una mera «valutazione di carattere essenzialmente sintomatico»)<sup>28</sup>.

---

BRICOLA, *La discrezionalità nel diritto penale*, Milano 1965, 402; MERZAGORA BETSOS, *Imputabilità e pericolosità sociale: un punto di vista criminologico e psicopatologico forense*, in *Verso un codice penale modello per l'Europa. Imputabilità e misure di sicurezza*, a cura di Manna, Padova 2002, 112; NOTARO, *Art. 203 c.p.*, in *Codice penale*, a cura di Padovani, Milano 2019, 1409 s.

<sup>23</sup> La Corte (Cass., Sez. II, 11 agosto 2020, in *Mass. Uff.*, n. 23797) continua correttamente evidenziando che tale fattibilità "tanto più doverosamente evitabile quanto più si rafforza il presupposto cognitivo, ossia l'analisi di tutto ciò che emerso sino al momento in cui la prognosi è richiesta (modalità dei fatti già realizzati, antecedenti causali, condotta di vita antecedente, fattori che possono aver inciso sulla determinazione ad agire, capacità di resistere alle spinte che muovono verso la trasgressione del precetto)".

<sup>24</sup> Cfr. MILETTO, *Le misure di prevenzione*, in *Le sostanze stupefacenti. Le misure di prevenzione*, a cura di Ambrosini, Miletto, in *Giurisprudenza sistematica di diritto penale*, a cura di Bricola, Zagrebelsky, Torino 1989, 145-146; NUVOLONE, voce *Misure di prevenzione e misure di sicurezza*, in *Enciclopedia del Diritto*, 1976, 632.

<sup>25</sup> KAISER, *Criminologia*, 1979, traduzione italiana a cura di Morselli, Blonck, Steiner, Milano 1985, 138.

<sup>26</sup> Cfr. SCHÖCH, *Kriminalprognose*, in KAISER-SCHÖCH, *Kriminologie/Jugendstrafrecht/Strafvollzug*, II, 1982, 85 ss.

<sup>27</sup> HASSEMER, *Einführung in die Grundlagen des Strafrechts*, München 1981, 244 ss.; così LO MONTE, *Gli interventi in tema di misure di prevenzione: il problema del congelamento di beni*, in *Le nuove norme di contrasto al terrorismo*, in *Commento al decreto -legge 27 luglio 2005, n. 144*, a cura di Dalia, Milano 2006, 434. Cfr. MANNA, *Il diritto delle misure di prevenzione*, in *Misure di prevenzione*, a cura di Furfaro, , Torino, 2013, 14 ss.

<sup>28</sup> Cass., 14 febbraio 1997, Nobile ed altri, in *Cass. pen.* 1997, 3171; Cass., 19 dicembre 1996, Di Muro, in *Cass. pen.* 1997, 2576; cfr. FRAGOLA, *Le misure di prevenzione*, Padova 1992, 14 ss., sui diversi metodi della prognosi criminale, e cioè quello intuitivo, dall'attendibilità discussa, e quello statistico, il quale utilizza come indici sintomatici di pericolosità valori negativi raccolti in tabelle prognostiche, che se pure presenta un maggiore fondamento scientifico, non è stato ancora sufficientemente sviluppato

Nel sistema nordamericano l'adozione di algoritmi predittivi trova la sua genesi all'indomani della riforma delle sentencing guidelines del 1984 che adottavano la teoria del just desert di stampo retributivo, ma soprattutto fondata sulla necessità di garantire l'*uguaglianza* di trattamento contro l'eccessiva discrezionalità giudiziaria volta a perseguire la finalità rieducativa, che si era registrata nel periodo precedente; ciò ha comportato un sovraffollamento carcerario legato all'espiazione massiva di pene detentive brevi, che ha fatto sorgere a sua volta la necessità che i giudici dovessero basare la decisione sulla quantità di pena - e, se del caso, quella sulla concessione di benefici premiali o di misure alternative alla detenzione - su evidenze statistiche (evidence-based practices). In tale direzione la valutazione attuariale del rischio di recidiva nel reato permette al giudice di assumere determinazioni più consapevoli e di scegliere la misura coercitiva o la quantità di pena più appropriata al caso di specie. In origine la valutazione del rischio era effettuata caso per caso dagli psicologi penitenziari, i quali si affidavano alle proprie conoscenze professionali e ai risultati del percorso riabilitativo svolto da condannato, ma tale sistema presentava il difetto di esprimere risultati difficilmente misurabili e confrontabili tra loro, oltre ad essere inutilizzabile nelle fasi giudiziarie che precedono l'esecuzione della pena. Tali problematiche hanno portato a sviluppare la evidence-based-practice attraverso l'utilizzo di strumenti predittivi sempre più sofisticati, che considerano l'interazione tra fattori di rischio statici e dinamici. Attualmente, le legislazioni di molti Stati prevedono che le Corti possano - e in molti casi debbano - considerare gli output forniti dall'algoritmo prima di assumere una determinata decisione<sup>29</sup>.

Si consideri che l'aspetto positivo degli algoritmi è che «progettano una procedura normativa che si sposta da un insieme di dati verso un *output* desiderato, escludendo le intuizioni soggettive e l'arbitrarietà dal processo. In questo

---

attraverso delle ricerche empiriche nei confronti di tutti i gruppi criminali o di gruppi ad espressa pericolosità; si denuncia, infatti, un uso del metodo intuitivo nella prassi e un scarso impegno della ricerca scientifica nel settore. Cfr. Cass., Sez. VI, 8 gennaio 2016, in *Mass. Uff.*, n. 8389.

<sup>29</sup> D'AGOSTINO, op. cit., 358 ss., che evidenzia che per le determinazioni in ordine all'applicazione delle misure cautelari personali e al rilascio su cauzione lo strumento più utilizzato è il Public Safety Assessment (PSA), un software che utilizza i dati di quasi 2 milioni di reati, commessi in 300 giurisdizioni degli Stati Uniti, per aiutare i giudici a decidere sulla libertà dell'indagato prima che questi sia rinviato a giudizio. In tempi recenti, gli algoritmi predittivi hanno assunto un ruolo centrale anche nella fase dibattimentale (trial), quando viene pronunciata una sentenza di condanna ed è necessario stabilire la specie e la quantità di pena da irrogare al caso concreto. Il primo Stato ad elaborare un proprio strumento di risk-assessment da utilizzare nella fase decisoria fu il Virginia nel 1994. Altri Stati preferirono utilizzare i prodotti commerciali allora esistenti, tra cui il Level of Service Inventory - Revised (LSI-R), in grado di combinare fattori statici e dinamici e di costruire un modello grafico per la determinazione il rischio di recidiva.

modo, rappresenta un modello matematico, che può essere gestito da un essere umano, anche in un procedimento penale, a condizione che si basi su una teoria convalidata e che tale teoria sia stata correttamente codificata nell'algoritmo. Questi due requisiti sono cruciali: [...], la possibilità di rivedere, discutere, sfidare algoritmi è una condizione di base per un equo procedimento penale, conforme ai diritti umani fondamentali»<sup>30</sup>.

In tale direzione si potrebbe sviluppare la tendenza che emerge nelle più recenti valutazioni di carattere c.d. attuariale, che costituiscono il presupposto teorico per l'utilizzo degli algoritmi predittivi, di affermare «una concezione *evidence-based* di valutazione del rischio individuale di commissione di un (nuovo) reato: una concezione, quindi, basata su riscontri oggettivi, destinata a soppiantare, o quanto meno integrare, le valutazioni intuitive dei giudici, tuttora ampiamente diffuse»<sup>31</sup>. Si valorizzerebbe, ad esempio, anche nella fase prognostica il carattere oggettivo e razionale della valutazione di pericolosità sociale, valorizzando i dati emergenti nella fase cognitiva del giudizio, basato, soprattutto in base alla più recente giurisprudenza in sede di applicazione delle misure di prevenzione sull'accertamento e la ricostruzione di fatti, o addirittura, di veri e propri precedenti giudiziari<sup>32</sup>.

Nell'ordinamento nordamericano si osserva, infatti, che continuare a fare affidamento esclusivamente sull'istinto e sull'esperienza del giudice o, comunque, di chi deve decidere, non è più sufficiente, o peggio potrebbe anche non essere etico, determinando una sorta di “sentencing malpractice” che produce raccomandazioni e sentenze che non sono né trasparenti né del tutto razionali<sup>33</sup>. Piuttosto, selezionare le possibili opzioni nella scelta della sanzione al fine

<sup>30</sup> QUATTROCOLO, *op. cit.*, 16.

<sup>31</sup> BASILE, *Intelligenza artificiale e diritto penale*, cit., 17; ZARA, *Tra il probabile e il certo. La valutazione dei rischi di violenza e di recidiva criminale*, in [www.penalecontemporaneo.it](http://www.penalecontemporaneo.it), con riferimento, in particolare, al lavoro di 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, 499 ss.

<sup>32</sup> Cfr. MAUGERI, *I destinatari delle misure di prevenzione tra irrazionali scelte criminogene e il principio di proporzione*, in *Indice Penale* 2017, 37 ss.; MAUGERI -PINTO DE ALBUQUERQUE, *La confisca di prevenzione nella tutela costituzionale multilivello: tra istanze di tassatività e ragionevolezza, se ne afferma la natura ripristinatoria (Corte Cost. n. 24/2019)*, in *Dir. pen. cont.*, 3, 2019, 97 ss.

<sup>33</sup> “Judges cannot impose appropriate sentences – those that will best help to protect public safety – without professionals conducting appropriate [evidence-based] violence risk assessment”. HART, *Evidence-Based Assessment of Risk for Sexual Violence*, 1 *Chapman J. Crim. Just.*, 2009, 143, 144. *It is irrational and unethical to make predictions based on legal or clinical experience alone, because research has shown this to be an invalid and inaccurate method for assessing risk*. Cfr. FAUST, ZISKIN, *The Expert Witness in Psychology and Psychiatry*, 241 *Science* 31, 33 (1988) (“If expertise is defined solely by accuracy, the actuarial method is the ‘expert.’”).

<sup>5</sup> Cfr. MEARES, *Three Objections t*

di individuare lo strumento che meglio potrebbe consentire di ridurre la possibilità di recidiva attraverso la rieducazione, o l'incapacitazione, o la deterrenza, è una questione scientifica che dovrebbe essere informata dalla scienza delle best practices (o meglio con un linguaggio familiare alla dottrina italiana alla migliore scienza ed esperienza), cioè "evidence-based practice", definite come: "pratiche professionali che sono supportate dalle "best research evidence", costituite da risultati scientifici relativi a strategie di intervento. . . derivate da ricerche clinicamente rilevanti. . . basate su revisioni sistematiche, dimensioni ragionevoli degli effetti, significatività statistica e clinica, e una massa di prove a sostegno. *Thus, the concept of evidence based practice in corrections refers to corrections practices that have been proven through scientific corrections research 'to work,' to reduce offender recidivism*"<sup>34</sup>. Si osserva, ancora, che la decisione basata sull'evidenza fa parte di un modello più ampio nella società contemporanea che implica l'uso della ricerca scientifica per migliorare la qualità del processo decisionale. "Come in medicina, psicologia, educazione, *managment* e altri campi, la scienza offre ora linee guida empiricamente fondate per l'esercizio della giustizia penale, nell'ambito di una tendenza graduale verso l'uso di pratiche basate sull'evidenza - *evidence-based practices* - nel diritto"<sup>35</sup>. In tal modo si dovrebbe anche individuare il trattamento individualizzato che meglio risponde alle reali necessità del reo, "*an individualized, evidence-based program tailored to each defendant's risk factors and needs*", per mettere in campo delle politiche - anche nella fase successiva alla detenzione - volte realmente a prevenire la recidiva<sup>36</sup> nella consapevolezza che "*[e]vidence-based sentencing merges punishment with rehabilitation. Imposing a sentence with appropriate conditions based on the defendant's individual risk to reoffend and need for treatment or programming does not diminish the prosecutor's role in advocating for appropriate sanctions; it makes the criminal justice system more effective*"<sup>37</sup>.

Si deve ricordare a tal proposito che «la valutazione *evidence-based* della pericolosità criminale presuppone la previa individuazione di una serie di *fattori*

---

<sup>34</sup> WARREN, *Evidence-based practices to reduce recidivism: implications for state judiciaries*, in *Crime & Justice Institute* 2007, 20; REDDING, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, in *Legal Studies Research Paper Series*, Paper No. 09-41, 2.

<sup>35</sup> Così HART, *op. cit.*, 145-46; REDDING, *op. cit.*, 2; Cfr. D'AGOSTINO, *op. cit.*, 359.

<sup>36</sup> HEILBRUN ET AL., *Risk-Assessment in Evidence-Based Sentencing: Context and Promising Uses*, 1 *Chapman J. Crim. Just.*, 127, 2009, 10: "We have two choices: we can continue the revolving door of recidivism or we can create policy to mandate evidence-based sentencing" (District Attorney for San Diego County).

<sup>37</sup> HEILBRUN ET AL., *op. cit.*, 135 ss.

*di rischio (o predittori)* direttamente coinvolti nel comportamento criminoso<sup>38</sup>, che possono consentire un *approccio di tipo attuariale (o statistico)* alla valutazione della pericolosità criminale. Attraverso, infatti, una loro combinazione, si possono predisporre delle “scale” che consentono l’attribuzione di un punteggio (*score*) al soggetto preso in esame<sup>39</sup>. Le “scale”, utilizzate per la valutazione attuariale della pericolosità criminale, si differenziano tra di loro, in base alla *popolazione* in relazione alla quale sono state elaborate<sup>40</sup>; alla *tipologia di reati implicati*<sup>41</sup>; alla *temporalizzazione del rischio* (immediato, o a medio o a lungo termine); al *contesto applicativo*<sup>42</sup>. I fattori di rischio, a loro volta, possono essere statici, non modificabili (ad esempio, il sesso e l’origine etnica); dinamici stabili, che sono modificabili grazie al trattamento terapeutico (ad esempio, il discontrollo degli impulsi); acuti, che cambiano rapidamente e sono associati ad una condizione facilitante la reazione violenta (ad esempio, l’uso di sostanze stupefacenti)<sup>43</sup>.

*3. I rischi connessi all’uso di algoritmi predittivi: discriminazioni e valutazioni generalizzanti.* La dottrina, tuttavia, correttamente evidenzia i rischi insiti nell’uso di simili meccanismi di risk assessment, già sperimentati negli Stati Uniti, e cioè che la scelta dei fattori di rischio – su cui si fondano gli algoritmi – non sia imparziale, ma basata su stereotipi scriminanti. Si evidenzia il rischio di c.d. *implicit bias*: da una parte laddove l’*input* non è completamente neutro, l’*output* dell’interrogazione rischia di essere influenzato da un pregiudizio, che può portare alla discriminazione di singoli individui o di gruppi sociali; dall’altra, l’algoritmo – che è concepito e interpretato da un umano – può banalmente riprodurre ingiustificati preconcetti sociali<sup>44</sup>. L’algoritmo po-

---

<sup>38</sup> Fattori che possono, tra l’altro, riguardare: l’età, il sesso, l’origine etnica, il livello di scolarizzazione, la situazione familiare e lavorativa, la posizione sociale, i precedenti penali, le precedenti esperienze carcerarie, i luoghi e le persone frequentati, la presenza di autori di reato nella cerchia familiare o nella rete di conoscenze, il luogo di residenza, le difficoltà di regolazione della rabbia e aggressività, il discontrollo degli impulsi, una storia di precedente violenza agita, una storia di ospedalizzazione, un pensiero pro-criminale, alcune variabili contestuali (quali, ad esempio, la mancanza di sostegno familiare e sociale), il consumo di sostanze stupefacenti o alcoliche, le psicopatie. Così BASILE, *Intelligenza artificiale e diritto penale*, cit., 17.

<sup>39</sup> Così BASILE, *Intelligenza artificiale e diritto penale*, cit., 17.

<sup>40</sup> Ad esempio, popolazione di adulti, di minori, di maschi, di femmine, di pazienti psichiatrico-forensi, di detenuti o ex-detenuti.

<sup>41</sup> Esistono scale generiche, cioè relative a tutti i reati, e scale specifiche, relative a singole tipologie di reati, come i reati sessuali o i reati violenti.

<sup>42</sup> Comunità civile, istituti di pena, centri di salute mentale, ospedali psichiatrico-giudiziari.

<sup>43</sup> Così BASILE, *Intelligenza artificiale e diritto penale*, cit., 17; cfr. QUATTROCOLO, *op. cit.*, 148 ss.

<sup>44</sup> QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., 6

trebbe rafforzare i c.d. stereotipi impliciti, fisiologicamente presenti nella persona che deve effettuare un giudizio, aumentando allora il rischio di un approccio proprio del diritto penale del tipo d'autore e del nemico; se «sostituiti algoritmi possono anticipare il comportamento di determinati soggetti, [...] il rischio è di cadere in una teoria lombrosiana di nuova generazione»<sup>45</sup>, il tutto in violazione, innanzitutto, del principio di uguaglianza<sup>46</sup>, del principio di offensività proprio del diritto penale del fatto (sancito dall'art. 25, comma 2, Cost.) e del principio di colpevolezza, correttamente inteso come colpevolezza per il fatto.

Ad esempio, nell'ordinamento nordamericano l'uso di algoritmi per la concessione della libertà su cauzione ha creato nuovi canali di discriminazione e favorito fenomeni di incarcerazione di massa per le fasce socialmente più deboli e povere<sup>47</sup>. Come evidenziato nel famoso caso Loomis<sup>48</sup>, ma anche in un più recente caso giudicato dalla Suprema Corte del Distretto della Columbia (relativa al SAVRY<sup>49</sup>)<sup>50</sup>, il problema consiste nel timore che

<sup>45</sup> ROMANO, *Diritto, robotica e teoria dei giochi: riflessioni su una sinergia*, in *Diritto e intelligenza artificiale*, a cura di Alpa, Pisa, 2020, 112; cfr. MANES, *op. cit.*, 690.

<sup>46</sup> MANES, *op. cit.*, p. 12 - 14; CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Dir. pubbl. comparato ed europeo*, 2019, 101 ss.

<sup>47</sup> ALEXANDER, *The Newest Jim Crow*, pubblicato in *The New York Times*, 8 novembre 2018. Invece di favorire “una nuova dimensione normativa basata sul calcolo, neutrale, e non sulla soggettività o il libero convincimento del giudice. La giustizia sarebbe, così, meglio assicurata dagli algoritmi (neutrali), piuttosto che dall'essere umano (soggetto a percezioni e a variabili soggettive ed imprevedibili)”, cfr. FRONZA, “Code is law”. note a margine del volume di antoine garapon e jean lassègue, *justice digitale. Révolution graphique et rupture anthropologique*, Puf, Paris, 2018, in *Dir. Pen. Cont.* 11 dicembre 2018; BURCHARD, *op. cit.*, 1932 ss.; MAYSON, *Bias In, Bias Out*, in *Yale law journal*, v. 128, n. 8, 2019, 2226 ss.

<sup>48</sup> Wisconsin Supreme Court, *State v. Loomis*, case 2015AP17-CR, Judgment July 13th 2016, in *130 Harvard L.R.*, 2017, 1530 ss. Cfr. QUATTROCOLO, *op. cit.*, 156 ss.; MALDONATO, *op. cit.*, 403 ss.; D'AGOSTINO, *op. cit.*, 362; CONTISSA, LASAGNI, SARTOR, *op. cit.*, 622 ss.; OCCHIUZZI, *Algoritmi predittivi: alcune riflessioni metodologiche*, in *Dir. pen. cont. Riv. Trim.* 2019, 394 ss.; OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models: Lessons from the Durham Constabulary HART model*, disponibile nel Sheffield Hallam University Research Archive (SHURA), <http://shura.shu.ac.uk/17462/>, 238.

<sup>49</sup> Structured Assessment of Violence Risk in Youth (SAVRY non è un *software*, ma un kit composto da guida e fogli di prova, per somministrare il test). È uno strumento professionale strutturato per la valutazione del rischio dinamico nel bambino, che pur utilizzando specifici dataset, è comunque comprovato da un parere clinico dell'esperto, finalizzato ai futuri sviluppi nella psiche del soggetto, cfr. ZARA, FARRINGTON, *Assessment of risk for juvenile compared with adult criminal onset: implications for policy, prevention and intervention*, in *Psychol Public Policy Law* 19(2), 2013, 235 ss. Si tratta, quindi, di un sistema che lascia un margine decisionale agli esperti ed è, quindi aperto e interpretabile, a differenza di COMPAS, cfr. S. Tola, M. Miron, E. Gómez, C. Castillo, *Why machine learning may lead to unfairness: evidence from risk assessment for Juvenile Justice in Catalonia*, Best Paper Award, International Conference on AI and Law, 2019; cfr. SAVIGNAC, *Tools to Identify and Assess the Risk of Offending Among Youth*, Published by *National Crime Prevention Centre (NCPC)*, 2010, che offre un quadro

l'algoritmo si basi e sia programmato sulla base di *biase date*<sup>51</sup>. «Per quanto il punteggio di rischio sia determinato processando i dati raccolti hic et nunc, rimane il fatto che il soggetto viene inquadrato in un profilo socio-criminale basato sul tasso di recidiva in consimili casu. L'obiezione principale mossa dagli studiosi americani riguarda l'inclusione, tra le variabili rilevanti ai fini della determinazione del livello di rischio, di fattori demografici, socioeconomici, familiari, che contribuiscono a caratterizzare come individui più pericolosi quelli appartenenti a determinate minoranze o classi sociali»<sup>52</sup>.

Il limite di tali strumenti, inoltre, è che comunque le prognosi si basano sui dati del passato, finendo per riprodurre nelle prognosi le discriminazioni del passato: «sia la previsione soggettiva che quella algoritmica guardano al passato come una guida al futuro e quindi proiettano in avanti le disuguaglianze del passato. Il vero problema, in altre parole, non è la metodologia algoritmica. Qualsiasi forma di previsione che si basi su dati sul passato produrrà disparità razziale se i dati passati mostrano che l'evento che aspiriamo a prevedere - la variabile presa di mira - si verifica con frequenza diseguale tra i gruppi razziali». «And if an algorithm's forecasts are correct at equal rates across racial lines, as were the COMPAS forecasts in Broward County, any disparity in prediction reflects disparity in the data»<sup>53</sup>. La pericolosità di un soggetto viene desunta esclusivamente dagli schemi comportamentali e dalle decisioni assunte in una determinata comunità del passato, in contrasto con il principio di individualizzazione del trattamento sanzionatorio, ex art. 27, c. 1 e 3, Cost., nonché, del canone di individualizzazione del trattamento cautelare, ricavabile dagli artt. 13 e 27, co. 2, Cost.<sup>54</sup>. A parte la considerazione che l'inserimento nell'algoritmo di informazioni relative ai precedenti penali, alle condanne già scontate, finisce per disconoscere la funzione rieducativa della pena, ad esempio quando si nega la libertà su cauzione in base al risk assessment calcolato da un algoritmo considerando anche tali precedenti; nell'ordinamento britannico si contesta che un simile algoritmo viola l'art. 4

---

sugli strumenti adoperati nel sistema canadese; CARDON, *Le Pouvoir des algorithmes des algorithmes*, Pouvoirs n°164 - La Datacratie - gennaio 2018, 65.

<sup>50</sup> Supreme Court of the District of Columbia, 25.3.2018, Judge Okun, commentato in QUATTROCOLO, *op. cit.*, 161 ss.

<sup>51</sup> QUATTROCOLO, *op. cit.*, 166 ss.; CHANDER, *The Racist algorithm*, 115 *Michigan Law Review*, 2017, 1023-1045; ANGWIN, "Machine Bias: There's Software Used Across The Country To Predict Future Criminals. And It's Biased Against Blacks", (Pro Publica, [www.propublica.org](http://www.propublica.org)); OCCHIUZZI, *op. cit.*, 396.

<sup>52</sup> D'AGOSTINO, *op. cit.*, 364.

<sup>53</sup> MAYSON, *op. cit.*, 2251.

<sup>54</sup> MANES, *op. cit.*, 14.

del *Rehabilitation of Offenders Act 1974*, “a rehabilitated person shall be treated ‘for all purposes in law’ as a person who has not committed the particular offence”<sup>55</sup>.

Più in generale emerge il timore che l’utilizzo di strumenti di IA contribuisca a una spinta in chiave securitaria del diritto penale quale strumento di controllo sociale, considerando le persone solo come potenziali rischi controllabili e costantemente valutabili (quindi, come soggetti pericolosi)<sup>56</sup>. Nell’ordinamento nordamericano si contesta, infatti, che simili strumenti sono funzionali all’affermazione della c.d. *New Penology*, in base alla quale lo scopo e responsabilità principale del diritto penale è “la gestione di “gruppi pericolosi” (management of “*dangerous groups*”)<sup>57</sup>, che sostituisce la c.d. “Old Penology”, il cui scopo principale è l’attribuzione di responsabilità per specifici atti criminali (“adjudication of guilt for specific criminal acts”). E un sistema di giustizia penale finalizzato all’incapacitazione del rischio finisce per perpetuare le discriminazioni del passato (e, quindi a perpetuare le ingiustizie razziali<sup>58</sup>); simili rischi si possono accentuare con l’utilizzo di algoritmi.

E, ancora, emergono i rischi connessi all’utilizzo di logiche elaborate su base collettiva (in riferimento a gruppi o classi di persone) che non tenga conto delle peculiarità della singola persona, anche perché per evitare il falso negativo tali programmi si basano proprio su un evento generico («i programmatori, ordinariamente, scelgono quale evento finale oggetto della predizione algoritmica, un evento generico, tale da incrementare il *data-set* di base»)<sup>59</sup>. Non solo, ma la dottrina nordamericana ha evidenziato che l’algoritmo presuppone inevitabilmente una serie di assunzioni implicite nel corso della progettazione, che «sottendono giudizi aventi ad oggetto sia questioni di politica cri-

---

<sup>55</sup> OSWALD, GRACE, URWIN, BARNES, *op. cit.*, 239 che ricorda come “ In *YA v London Borough of Hammersmith and Fulham* [2016] EWHC 1850, a local authority acted unlawfully in taking spent convictions into account in refusing the claimant a place in social housing (since the decision related, the court must have felt, to a ‘legal purpose’); while in *N v Governor of HMP Dartmoor* [2001] EWHC Admin 93, a prison governor lawfully shared with social services particular details of spent convictions of a prisoner re-entering the community on license, since public protection is not a ‘legal purpose’ in the view of the court in that case, but a matter of public policy”.

<sup>56</sup> BURCHARD, *op. cit.*, 1909 ss.

<sup>57</sup> MALCOLM FEELEY, SIMON, *The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications*, in 30 *Criminology* 1992, 456; MAYSON, *op. cit.*, 2281 s.

<sup>58</sup> MALCOLM FEELEY, SIMON, *op. cit.*, 456.

<sup>59</sup> MALDONATO, *op. cit.*, 405, “La dottrina d’oltreoceano che più da vicino si è occupata del tema ha evidenziato che gli sviluppatori degli strumenti in esame compiono inevitabilmente una serie di assunzioni implicite nel corso della progettazione; come si vedrà, tali assunzioni sottendono giudizi aventi ad oggetto sia questioni di politica criminale sia questioni relative alle circostanze valutabili in sede di giudizio (ad es. la personalità del reo, il suo carattere, le sue inclinazioni)”.

minale sia questioni relative alle circostanze valutabili in sede di giudizio (ad es. la personalità del reo, il suo carattere, le sue inclinazioni)»<sup>60</sup>.

Gli stessi problemi del giudizio intuitivo riaffiorano, ma esasperati, perché «la discrezionalità soggettiva si riscontra lungo tutto l’arco che procede dalla raccolta dei dati ai risultati definitivi basati sui medesimi...<sup>61</sup>. [...] La stessa costruzione degli algoritmi e le loro finalità volte all’estrazione (*data mining*), al confronto (*data matching*) o alla profilazione (*data profiling*), nonché la preferenza per uno piuttosto di un altro tra essi, non sono certamente esenti da opzioni inevitabilmente personali nella formazione e nella scelta sia dei vari gruppi (i *cluster*) in cui sono raccolti e suddivisi gli elementi sia del peso che si intende attribuire loro: l’algoritmo “è strutturalmente condizionato dal sistema di valori e dalle intenzioni di chi ne commissiona la creazione e/o di chi lo elabora<sup>62</sup>”; ma ai valori individuati dal programma – in una società democratica e pluralista – sarà sempre più difficile assegnare convenzionalmente un punteggio fisso ‘buono per tutti’ e quindi oggettivo (come invece potrebbe fare un algoritmo [deterministico])”<sup>63</sup>. Nell’algoritmo si esasperano, del resto, le problematiche e i limiti degli stessi meccanismi di valutazione del rischio e, in particolare, di accertamento della pericolosità; la previsione di pericolosità, in particolare, si presenta come assolutamente problematica perché non solo inesatta, ma perché esprime un giudizio di valore che deve essere adeguatamente riservato al decisore legale<sup>64</sup>.

Nell’uso di algoritmi emerge, allora, il rischio di adottare logiche proprie del diritto penale del nemico, laddove la scelta dei fattori di rischio sia *biased*, fondata su pregiudizi di sesso, di razza, di religione, ..., o del tipo di criminalità, come ad esempio quella di carattere sessuale. Nella prassi statunitense si ritiene che si rischia di adottare la teoria dell’inabilitazione selettiva nell’adozione degli *input* e nell’esito decisorio (pena più severa per un individuo ritenuto socialmente pericoloso per l’appartenenza ad una “categoria”)<sup>65</sup>;

<sup>60</sup> MALDONATO, *op. cit.*, 405.

<sup>61</sup> Occorre non solo scegliere il genere di elementi da archiviare nella memoria elettronica, bensì anche stabilire i loro requisiti e il loro numero minimo ritenuti qualitativamente e quantitativamente sufficienti per essere significativi.

<sup>62</sup> SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020, 614.

<sup>63</sup> Così UBERTIS, *op. cit.*, 5; PARODI - SELLAROLI, *op. cit.*, 67 ss.; DI GIOVINE, *op. cit.*, 965.

<sup>64</sup> SCURICH, *The case against categorical risk estimates*, in *Behavioral Science Law* 2018, 1 ss. - 559

<sup>65</sup> Cf. MATHIESEN, *Selective Incapacitation Revisited*, in *Law and Human Behavior*, 1998, 22, 4, 455; D’AGOSTINO, *op. cit.*, 357 che ricorda che “la teoria si fondava sull’assunto che i criminali professionali o per tendenza – responsabili dei delitti più gravi – possono essere facilmente individuati partendo da alcune caratteristiche note, come la loro storia personale e criminale. Tuttavia, la scelta di punire i cri-

tale teoria viene spesso richiamata nel dibattito attuale per mettere in dubbio l'attendibilità degli algoritmi predittivi e dell'utilizzo di generalizzazioni empiriche basate sull'appartenenza dell'individuo a un gruppo<sup>66</sup>. "La negazione della qualità di persona all'*outsider* sociale riflette del resto i paradigmi del lessico attuariale: le statistiche sui fattori di rischio e sui tassi di criminalità riducono il singolo, nella propria unicità e complessità personologica, ad un novero di parametri di rischio, deterministicamente prevedibile nel proprio agire, cui si nega la contingenza comportamentale e la libertà di scelta, sostituite dalle certezze deterministiche dei calcoli fattoriali"<sup>67</sup>. Il soggetto è così ridotto a mero «*data-constructed bearer of the characteristics of danger*», per il quale «*there is no way back to becoming a rational moral agent; the route from the fortress to the wilderness is one-way*»<sup>68</sup>.

Anche sotto il profilo del trattamento emerge il rischio che alla base del modello dominante di riabilitazione (trattamento correzionale) basato sul Risk-Need-Responsivity si ponga non l'idea di un soggetto responsabile delle sue azioni, ma piuttosto una visione antropologica conflittuale e deterministica, che "vede il trasgressore come portatore di rischi e come oggetto passivo dell'intervento, proprio come la macchina da riparare è vista dall'ingegnere"<sup>69</sup>. La valutazione di pericolosità sociale non sarebbe più individualizzata tenendo conto della base cognitiva di tale giudizio nel caso concreto, ma per basarsi su valutazioni standardizzate finendo per presumere la pericolosità, presun-

---

minali non per il fatto già commesso, ma per quello che avrebbero potuto commettere in futuro, si scontrava con la tesi di coloro che, sulla base di evidenze statistiche, dimostravano che i crimini attesi potevano in concreto non essere mai commessi" (MATHIESEN, *op. cit.*, 455). "La prevenzione del crimine mediante la previsione ha accompagnato il sistema di giustizia criminale degli Stati Uniti a partire dagli anni '20. Durante gli anni Sessanta e i primi anni settanta, gli studi si sono concentrati principalmente sulla ricerca degli indici di pericolosità che confermassero l'attitudine del soggetto alla commissione di crimini violenti. Pur di fronte alle difficoltà di stabilire in modo oggettivo gli indici rivelatori della pericolosità attuale, i fautori della teoria in commento proponevano, seguendo un approccio di tipo utilitaristico, di punire alcuni individui più severamente sulla base del solo giudizio prognostico positivo di recidiva nel reato. La scarsa affidabilità scientifica del metodo ha contribuito a rendere la teoria dell'inabilitazione selettiva un retaggio del passato; residuano tuttavia alcuni istituti che sembrano rievocarla. Numerosi Stati hanno introdotto statuti autonomi di disciplina per i criminali seriali, istituito presso le Procure reparti specializzati in procedimenti contro criminali professionali, e imposto ai giudici di tener conto dei precedenti penali, della stabilità lavorativa e di altri dati personali" (così D'AGOSTINO, *op. cit.*, 355; COHEN, *Incapacitation as a Strategy for Crime Control: Possibilities and Pitfalls*, in *Crime and Justice*, University of Chicago Journals - Crime and Justice, 1983, 5, I, 12).

<sup>66</sup> D'AGOSTINO, *op. cit.*, 358.

<sup>67</sup> Così RESTA, *Nemici e criminali. Le logiche del controllo*, in *Ind. Pen.* 2006, 193 s.

<sup>68</sup> HUDSON, *Justice in the Risk Society*, London-Thousand Oaks-New Delhi 2003, 76.

<sup>69</sup> WALGRAVE, WARD, ZINSSTAG, *When restorative justice meets the Good Lives Model: Contributing to a criminology of trust*, in *Eur. Journ. of Criminology*, 2019, 3; PEREIRA, *op. cit.*

zioni bandite già in epoca risalente dalla Corte costituzionale<sup>70</sup>, in contrasto con il principio di colpevolezza e la presunzione d'innocenza, nonché il principio di uguaglianza e di proporzione, basandosi su giudizi discriminatori.

Anche nell'ordinamento nordamericano nel 2014 il Department of Justice's Criminal Division ha espresso una posizione piuttosto scettica nei confronti degli algoritmi predittivi in una lettera al presidente della U.S. Sentencing Commission, intitolata "The Promise and Danger of Data Analytics in Sentencing and Corrections Policy", evidenziando il rischio di violare la clausola costituzionale di pari protezione e mettendo in guardia i legislatori nazionali sui possibili effetti discriminatori del loro utilizzo su individui provenienti da classi sociali disagiate<sup>71</sup>. Nello stesso anno il procuratore generale degli Stati Uniti, Eric Holder, durante una manifestazione della National Association of Criminal Defense Lawyers, ha denunciato il rischio che strumenti come i *software* di valutazione del rischio possano «esacerbare disparità ingiustificate e ingiuste», già presenti nel sistema penale americano<sup>72</sup>.

La "Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relative ambienti" del 2018<sup>73</sup>, sopra citata, ha precisato in relazione ai procedimenti penali che «anche se non sono specificamente progettati per essere discriminatori, l'uso di algoritmi basati sull'IA [...] ha mostrato il rischio di favorire la rinascita di teorie deterministiche a scapito delle teorie dell'individualizzazione della pena» (p. 48); si afferma, quindi, il divieto di discriminazioni nei confronti di gruppi o individui nell'uso di processi computazionali utilizzati in sede giudiziaria. Nella proposta di Regolamento (Artificial Intelligence Act) 21 aprile 2021, si evidenzia che se il sistema di IA non è addestrato con dati di alta qualità, non soddisfa i necessari requisiti di *accuratezza o robustezza*, o non è adeguatamente progettato e testato prima di essere immesso sul mercato o altrimenti messo in servizio, *può individuare le persone in modo discriminatorio o comunque scorretto o ingiusto*.

---

<sup>70</sup> Corte cost., n. 1 del 1971, nota di VASSALLI, con la quale la Corte ha dichiarato illegittima la presunzione di pericolosità del minore non imputabile; Corte cost., n. 139 del 1982, con nota di MUSCO, in *Riv. it. dir. pen. proc. pen.* 1982, 1585, con la quale la Corte ha dichiarato illegittima la presunzione di pericolosità del prosciolto per infermità di mente (art. 222, comma 1, c.p.); Corte cost., 28 luglio 1983, n. 249, in *Riv. it. dir. pen. proc. pen.*, 1984, 460, con nota di Giuri, con la quale la Corte ha dichiarato illegittima la presunzione di pericolosità del seminfermo di mente (art. 219 c.p.).

<sup>71</sup> KEHL, GUO, KESSLER, *Algorithms in the Criminal Justice System: Assessing the use of Risk Assessments in Sentencing*, *Responsive Communities Initiative, Berkman Klein Center for Internet and Society (Harvard Law School, <https://dash.harvard.edu>)*, 2017, 16

<sup>72</sup> Cfr. QUATTROCOLO, *op. cit.*, 170.

<sup>73</sup> Cfr. UBERTIS, *op. cit.*, 7.

Il Decreto legislativo 18 maggio 2018 n. 51, introdotto in attuazione della Direttiva 2016/680/UE<sup>74</sup>, vieta all'art. 8, c. 4 «la profilazione finalizzata alla discriminazione di persone fisiche sulla base di categorie particolari di dati personali di cui all'articolo 9 del regolamento UE».

4. *La mancanza di trasparenza.* Un altro problema che emerge nell'utilizzo di algoritmi è dato dalla mancanza di *trasparenza* del metodo utilizzato dall'algoritmo predittivo, - del modo in cui vengono a connettersi le informazioni già presenti nel calcolo e quelle relative al singolo caso -, che è ordinariamente un algoritmo proprietario, coperto cioè da diritto d'autore e, quindi, non conoscibile dalle Corti, né controllabile dalla difesa, con la conseguenza che viene a mancare il contraddittorio sull'ammissibilità dell'utilizzo dello strumento e sulle sue risultanze con violazione del principio del giusto processo<sup>75</sup>, della presunzione d'innocenza e dei diritti della difesa.

Nel caso *Loomis* la difesa lamentava giustamente una generale carenza di informazioni sia sui parametri per la determinazione del rischio all'interno della teoria psico-criminologica applicata, sia sul codice sorgente che regola il *software*, con conseguente, grave limitazione alla possibilità di falsificazione; la Corte Suprema del Wisconsin ha sempre respinto questi argomenti<sup>76</sup>. La "*black box decision*" viene paragonata ad una decisione priva di motivazione o con motivazione meramente apparente (in contrasto con l'art. 24 Cost. e con l'art. 111/6 Cost.), o ad una decisione assunta in un contesto di asimmetria informativa (*knowledge impairment*) in contrasto con il principio dell'*equality of arms* che si esprime, in base alla giurisprudenza della Corte EDU, nella facoltà di contestare e criticare le prove contrarie<sup>77</sup>.

<sup>74</sup> Relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio. (18G00080) (GU Serie Generale n.119 del 24-05-2018).

<sup>75</sup> MALDONATO, *op. cit.*, 409.

<sup>76</sup> QUATTROCOLO, *op. cit.*, 166.

<sup>77</sup> MANES, *op. cit.*, 691. Paragonando la "*black box decision*" a una perizia si può ricordare che, come evidenziato dalle S.U., Cass., Sez. un., 2 aprile 2019, Pavan, in *questa Rivista*, "ad avviso della Corte EDU, il principio della parità delle armi comporta, relativamente ai "testimoni esperti": - il diritto di partecipare alla formazione della prova in tutte le sue fasi; l'audizione di persone che possono essere chiamate, a qualsiasi titolo, dalla parte che vi abbia interesse a sostenere la propria tesi (Doorson c. Paesi Bassi §§ 81-82; la possibilità di nominare un "contro-esperto": Bónisch c. Austria); - la possibilità di partecipare all'esame delle persone (*cross examination*) sentite dal perito; il diritto di prendere cognizione tempestivamente dei documenti utilizzati dal perito al fine di poterli valutare (Mantovanelli c. Francia); - il diritto di esaminare direttamente i periti (Matytsina c. Russia; Poletan e Azirovik c. Macedonia); la violazione del diritto di difesa e, quindi del giusto processo nel caso in cui: il giudice negli

Dell'algoritmo, allora, occorre garantire per essere validamente utilizzato in sede penale proprio la trasparenza che lo rende controllabile: «In questo senso, un'efficace trasparenza dipende, in primo luogo, dalla precisione della teoria scientifica alla base e, in secondo luogo, dalla chiarezza del linguaggio utilizzato per tradurlo in una formula matematica [...] un linguaggio matematico chiaro permette un *ex post reviewer* per capire come si è evoluto il processo dagli *input* agli *output*»<sup>78</sup>. Per evitare effetti di sovrastima e/o rischi di “falsi positivi”, e per consentire alla difesa di verificare la scientificità e l'accuratezza di un “enigmatico database” o di un dato generato da un determinato processo computazionale si dovrebbe ammettere nei confronti dell'algoritmo la valutazione peritale, – non diversamente da ogni acquisizione scientifica che entri nel processo penale –, e comunque la sua fondatezza empirica dovrebbe essere valutata in contraddittorio, nel rispetto dei diritti della difesa<sup>79</sup>.

In materia è stato correttamente evidenziato che le costituzioni nazionali e le carte dei diritti sovranazionali impongono di far prevalere l'interesse dell'imputato o comunque del soggetto interessato (indiziato) a un processo equo, che è poi un pubblico interesse, sull'interesse privato di carattere economico a tenere segreto l'algoritmo<sup>80</sup>. Vengono in gioco i principi fondamentali del giusto processo e del diritto della difesa, sanciti innanzitutto dall'art. 6 della C.E.D.U. e dagli artt. 47 e 48 della Carta di Nizza.

In tale direzione nella “Carta etica europea” del 2018 il quarto principio enunciato è quello di trasparenza, imparzialità e fairness; si raccomanda *l'accessibilità, la comprensibilità e la verificabilità esterna dei processi computazionali* utilizzati per l'analisi dei dati giudiziari, emergendo la preoccupazione in merito al complesso rapporto tra protezione della proprietà intellettuale e del *trade secret* e la necessità di osservare, capire, criticare i processi com-

---

immotivatamente il diritto della parte di chiedere una controperizia, quando il perito, in sede dibattimentale, muti radicalmente l'opinione che aveva espresso nella relazione scritta; il Collegio peritale sia formato in maggioranza da professionisti dipendenti dalla struttura chiamata a rispondere dei danni subiti dalla persona offesa (GB. c. Francia). La giurisprudenza europea, conferma, quindi, che le regole del “giusto processo” si applicano anche ai “testimoni esperti”, laddove costoro, come i testimoni “comuni”, abbiano un “peso significativo” nella decisione assunta dal giudice (Poletan e Azirovik c. Macedonia)».

<sup>78</sup> Tuttavia, per soddisfare il bisogno di giustificazione, la teoria scientifica di base deve essere abbastanza valida da fornire una relazione di causalità tra l'insieme dei dati immessi e i risultati. Nell'ambito di questo studio, ci riferiamo al concetto di “validità” il senso fornito dalla Corte Suprema degli Stati Uniti nel caso menzionato di Daubert.”, così QUATTROCOLO, *op. cit.*, 17 – cfr. 163.

<sup>79</sup> MANES, *op. cit.*, 16.

<sup>80</sup> QUATTROCOLO, *op. cit.*, 177.

putazionali utilizzati; principio valido per tutti i settori della pubblica amministrazione, ma a maggior ragione in quello della giustizia e, nell'ambito di questa, nei confronti del processo penale, in cui sono in gioco i più elevati beni giuridici, come la libertà personale e la reputazione<sup>81</sup>. Strettamente connesso al principio di trasparenza è anche il terzo principio riconosciuto nella Carta, e cioè quello di qualità e sicurezza, che raccomanda, *con riguardo all'analisi dei dati e delle decisioni giudiziarie, l'uso di fonti certificate e dati intangibili, attraverso modelli concepiti in modo multidisciplinare, in un ambiente tecnologico sicuro*; “questo principio che, secondo i redattori, sottende la creazione di piccoli gruppi interdisciplinari di lavoro, ispirati alla eccellenza professionale e al rispetto dei principi etici in questione, permea tutti i punti del documento”<sup>82</sup>. Anche nella proposta di Regolamento (Artificial Intelligence Act) del 2021, si evidenzia (considerando n. 38) che l'esercizio di importanti diritti procedurali fondamentali, come il diritto a un ricorso effettivo e a un giusto processo, nonché il diritto alla difesa e alla presunzione di innocenza, potrebbero essere compromessi laddove tali sistemi di IA, utilizzati nel sistema di giustizia penale, non siano sufficientemente *trasparenti, spiegabili e documentati*.

Alcuni studiosi considerano la trasparenza l'antidoto al carattere discriminatorio dell' algoritmo. Si parte dal presupposto che non sia possibile garantire la neutralità dell' algoritmo in considerazione del *rischio intrinseco* proprio della decisione algoritmica in quanto ritengono che la necessità di servirsi di dati che non possono che essere sempre parziali conduce a una situazione di *bias inevitabile*, nel senso che sarebbe impossibile fornire una rappresentazione neutra dell'informazione iniziale; non si ritiene possibile “esigere la neutralità da parte di un sistema progettato per scegliere, filtrare o ordinare le informazioni secondo certi principi e concepito proprio al fine di ridurre l'incertezza in un universo in cui l'abbondanza di dati non consente di scegliere”<sup>83</sup>. Allora, si preferisce, a fronte di una *vana neutralità*, esigere la *lealtà dell'algoritmo*<sup>84</sup> e cioè la sua *trasparenza*, rendendo conoscibili i criteri che presiedono la decisione algoritmica e il funzionamento del sistema, consentendo l'accesso alle fasi del ragionamento seguite dal sistema per assumere una determinata decisione<sup>85</sup>; il tutto al fine di evitare che la parzialità della conoscenza su cui si basa

<sup>81</sup> QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., 7.

<sup>82</sup> *Idem*, 6.

<sup>83</sup> CARDON, *Le Pouvoir des algorithmes*, cit., 65.

<sup>84</sup> *Idem*, 66.

<sup>85</sup> È la prospettiva sostenuta anche da FIORIGLIO, *La “dittatura” dell'algoritmo*, cit., 137.

la decisione possa celare una discriminazione non identificabile<sup>86</sup>. Si tratta della caratteristica dell'*esplicabilità* (*explicability*)<sup>87</sup>, che deve connotare i sistemi di IA intesi come sistemi che rispondono al requisito della trasparenza. Il sistema di IA risponderà al criterio della esplicabilità se il produttore è in grado di anticipare gli effetti cui può condurre l'utilizzo del sistema, in modo tale che l'utilizzatore possa per tempo identificarli e, se del caso, segnalare eventuali anomalie<sup>88</sup>. Naturalmente, il livello di trasparenza richiesto aumenta allorché l'algoritmo decisionale sia in grado di incidere su valori o diritti fondamentali<sup>89</sup>.

Quando, però, la discriminazione è riconducibile ai dati e l'algoritmo replica le discriminazioni in essi contenuti a causa del metodo statistico, la *transparency by design* o *explicability* potrebbe non essere sufficiente. Il sistema, infatti, non presenterebbe errori nel modo in cui funziona e la discriminazione sarebbe il riflesso delle disuguaglianze che connotano la società<sup>90</sup>. In questo caso, il correttivo potrebbe essere individuato non tanto nella *transparency by design*, ma nella *transparency degli input e degli output*<sup>91</sup>.

Questo sistema di controllo dell'output potrebbe anche essere impiegato quando non si è in grado di spiegare come l'algoritmo sia pervenuto a una data decisione ovvero quando l'algoritmo si sia modificato a seguito dell'autoapprendimento<sup>92</sup>.

In materia penale il principio di trasparenza assume un significato particolare in quanto si connette alla necessità di garantire il principio fondamentale della *pubblicità del processo decisionale* e, in particolare, in sede di valutazione

---

<sup>86</sup> Così GIUCA, *op. cit.*, 39

<sup>87</sup> "Tale principio implica che i processi devono essere trasparenti, le capacità e lo scopo dei sistemi di IA devono essere comunicati apertamente e le decisioni, per quanto possibile, devono poter essere spiegate a coloro che ne sono direttamente o indirettamente interessati", ed è di uno dei principi etici elaborati dal Gruppo di esperti ad alto livello sull'intelligenza artificiale nel documento *Orientamenti Etici per un'IA affidabile*, 14. I principi o imperativi etici sono quattro: i) il rispetto dell'autonomia umana, ii) la prevenzione dei danni, iii) l'equità, iv) l'esplicabilità.

<sup>88</sup> CARDON, *op. cit.*, 70.

<sup>89</sup> Come sottolineato dal Gruppo di esperti ad alto livello sull'intelligenza artificiale nel documento *Orientamenti Etici per un'IA affidabile*, cit.

<sup>90</sup> CHANDER, *op. cit.*, 1039.

<sup>91</sup> *Idem*, 1039. Cfr. anche *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, Committee of experts on internet intermediaries (MSI-NET), Council of Europe, March 2018, ove, in tema di trasparenza dell'algoritmo, si precisa che: "there may be a possibility of demanding that key subsets of information about the algorithms be provided to the public, for example which variables are in use, which goals the algorithms are being optimised for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm", 38.

<sup>92</sup> CHANDER, *op. cit.*, 1039.

della prova, di garantire che la motivazione della sentenza dia conto della valutazione di attendibilità operata dal giudice rispetto a ciascuna prova<sup>93</sup>.

A questo scopo, utile opzione collaterale è individuata dalla Carta etica europea nella creazione di autorità indipendenti che possano verificare e certificare *a priori*, e periodicamente, i modelli impiegati nei servizi della giustizia.

Ferma restando l'imprescindibile necessità della trasparenza, non manca chi evidenzia la difficoltà di garantire effettivamente tale trasparenza e la comprensione del funzionamento dell'algoritmo cui tale trasparenza è funzionale, nonché il rischio di opacità che il sistema mantiene in quanto "sta diventando sempre più difficile spiegare a scienziati non informatici e non statistici in che modo un machine learning forecasting model arriva ai suoi risultati", ne consegue che "il potenziale per incomprensioni e persino travisamenti intenzionali è vasto"<sup>94</sup>: "anche là dove il *reverse engineering* sia possibile, la comprensione del modello rimane questione limitata ai soli esperti, con esclusione degli effettivi destinatari della 'decisione automatizzata'<sup>95</sup>. Si teme che la 'trasparenza algoritmica' non sia in grado di fornire al giudice, ai destinatari della decisione e all'opinione pubblica l'effettiva comprensione del processo che ha portato a generare la prova digitale, e, ancor più importante in sede penale, il giudizio sulla sua attendibilità<sup>96</sup>.

5. *L'adozione del modello "Daubert" per verificare la scientificità della teoria psico-criminologica e del software.* Uno dei rischi connessi all'utilizzo di algoritmi predittivi è legato, come evidenziato, all'utilizzo di logiche elaborate su base collettiva (in riferimento a gruppi o classi di persone) che non tenga conto delle peculiarità della singola persona; il soggetto viene inquadrato in un profilo socio-criminale basato sul tasso di recidiva in *consimili casu* finendo per perpetuare le discriminazioni del passato. Un soggetto appartenente ad una categoria "a rischio" è ritenuto più pericoloso sulla base di mere generalizzazioni (*group-based generalizations*)<sup>97</sup>.

In materia emergono dei rischi simili a quelli emergenti nell'uso di leggi di carattere epidemiologico per l'accertamento del nesso causale, che proprio per tali limiti non possono fondare un giudizio di causalità ai fini della re-

<sup>93</sup> QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., 8.

<sup>94</sup> Cfr. QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., 8; KOENE - WEBB - PATEL, *First UnBias Stakeholders workshop*, 11 in <https://unbias.wp.horizon.ac.uk/dissemination/publications/>

<sup>95</sup> OSWALD, GRACE, URWIN, BARNES, *op. cit.*, nota 74. Cfr. KOENE - WEBB - PATEL, *op. cit.*, 234; QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., 8.

<sup>96</sup> QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., 8.

<sup>97</sup> D'AGOSTINO, *op. cit.*, 365.

sponsabilità penale, in quanto si sostituirebbe al necessario giudizio individualizzante che tiene conto delle peculiarità del caso concreto (pur cercando una spiegazione dei fatti, sussumendo il fatto concreto in una legge scientifica valida, di carattere generalizzante), un mero giudizio generalizzante di carattere statistico probabilistico<sup>98</sup>.

Tanto è vero che la dottrina (e anche la Suprema Corte del District of Columbia)<sup>99</sup> ritiene che si dovrebbero adottare i parametri stabiliti nella sentenza Daubert della Corte Suprema americana<sup>100</sup> per stabilire la validità di una legge scientifica (nell'applicazione della Rule 702 dell'US Federal Rules of Criminal Procedure), anche al fine di determinare la validità, innanzitutto, della teoria scientifica di carattere psicocriminologico che dovrebbe essere alla base dell'algoritmo e poi dello stesso algoritmo, come traduzione in termini computazionali della teoria<sup>101</sup>. In termini simili autorevole dottrina italiana ha suggerito l'uso dei criteri della sentenza Daubert per consentire al giudice di valutare la validità di una legge scientifica da utilizzare in sede penale, in particolare ai fini dell'accertamento della causalità<sup>102</sup>, considerando che la Suprema Corte italiana pur riconoscendo nella ben nota sentenza Bonetti, sul disastro di Stava, che le leggi scientifiche dovessero «ricevere conferma mediante il ricorso a metodi di prova razionali e controllabili», non specificava quali fossero questi metodi<sup>103</sup>; suggerimento in qualche modo seguito dalla successiva giurisprudenza della Suprema Corte<sup>104</sup>.

<sup>98</sup> Cfr. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Milano 2001, 228 ss..

<sup>99</sup> Supreme Court of the District of Columbia, 25.3.2018, Judge Okun, cit.

<sup>100</sup> *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). La sentenza è riportata per esteso in STELLA, *Leggi scientifiche e spiegazione causale nel diritto penale*, Milano 2000, 2 ed., 424 ss. Nella sentenza *Kumho Tire Co. v. Carmichael*, 119 S.Ct. 1167 (1999), l'applicazione del "Daubert test" è stata estesa al settore tecnico. Cfr. TONINI, *La Cassazione accoglie i criteri Daubert sulla prova scientifica. Riflessi sulla verifica delle massime di esperienza*, in *DPP*, 2011, 1345 ss.; BLAIOTTA, CARLIZZI, *Liberò convincimento, ragionevole dubbio e prova scientifica*, in a cura di Canzio, Luparia, *Prova scientifica e processo penale*, Padova 2018, 367 ss.; CARLIZZI, *Iudex peritus peritorum*, in *Dir. pen. cont. Riv. trim.*, 2017, n. 2, 28 ss.

<sup>101</sup> Cfr. QUATTROCOLO, *op. cit.*, 156.

<sup>102</sup> STELLA, *op. cit.*, 315 ss.

<sup>103</sup> Cass., 6 dicembre 1990, Bonetti, in *Il Foro it.* 1992, II, 35 ss.

<sup>104</sup> Cass., Sez. un., 10 luglio 2002, Franzese, in *Mass. Uff.*, n. 222138-222139; Cass., Sez. IV, 17 settembre 2010, Cozzini, *ivi*, n. 248943-248944; Cass., Sez. IV, 29 gennaio 2013, Cantore, *ivi*, n. 255105 che, come evidenziato nella sentenza delle Sez. un. 2 aprile 2019, Pavan, cit., "hanno segnato un punto di svolta nell'approccio alla problematica della valutazione della prova scientifica essendosi introdotto nelle aule giudiziarie il principio della falsificabilità della prova (secondo il quale una tesi scientifica non può mai essere provata in modo certo, ma può essere solo falsificata) e, quindi, l'abbandono del metodo induttivo per enumerazione (secondo il quale una tesi scientifica può ritenersi convalidata quando un determinato risultato resta confermato dalla ripetizione di un numero finito di casi identici)"; cfr.

Per garantire allora l'accessibilità e la falsificazione del "test algoritmico", nonché l'affidabilità del metodo scientifico predittivo, alla base del modello computazionale<sup>105</sup> e cioè la correttezza della traduzione della teoria psico-criminologica (metodo scientifico predittivo) in linguaggio algoritmico, si deve pretendere e accertare a monte la validità della teoria scientifica adottata nel calcolo della pericolosità e la validità del modello di *software* utilizzato. Se si applicassero i criteri utilizzati nella sentenza Daubert per affermare la validità di una legge scientifica alla luce della combinazione dei diversi criteri elaborati in dottrina, si dovrebbe verificare: (a) se la tecnica o la teoria dell'esperto è stata (o può essere) testata (cioè se la teoria dell'esperto può essere empiricamente verificata, o se è a malapena un approccio soggettivo, apodittico, che non può essere ragionevolmente valutato per l'affidabilità; la corrente metodologica scientifica «si basa sulla formulazione di ipotesi sul controllo delle stesse, per vedere se possano essere falsificate»<sup>106</sup>); (b) se la tecnica o la teoria è stata oggetto di pubblicazione e revisione tra pari; (c) il tasso di errore noto (o potenziale) nell'applicazione della tecnica o della teoria; (d) l'esistenza e il mantenimento di standard e controlli; e (e) se la tecnica o la teoria sia accolta dalla relativa comunità scientifica<sup>107</sup>. La Corte utilizza, insomma, i criteri proposti da contrapposte teorie della conoscenza, per valutare se l'ipotesi scientifica rispetti il metodo scientifico: la conferma dell'ipotesi secondo il metodo induttivistico (Hempel), la sottoposizione a tentativi di falsificazione (Popper), il consenso generale della comunità scientifica (Kuhn, il "consenso generale nei periodi di scienza normale")<sup>108</sup>. Il tutto evidenziando che il giudice «eserciti

---

Cass., Sez. IV, 17 settembre 2010, Cozzini, § 14 sul ruolo del giudice come *peritus peritorum*, nel senso che il giudice deve assolvere un ruolo critico e non di mero spettatore passivo di fronte allo scenario prospettato dal sapere scientifico, elevandosi a custode della scientificità del metodo; e in tale prospettiva la Cassazione individua taluni indici sintomatici della affidabilità e integrità dell'apporto tecnico-scientifico: 1) l'apprezzamento in ordine alla qualifica professionale e indipendenza dell'esperto (perito); 2) la valutazione dello stato complessivo delle conoscenze accreditate; 3) il grado di consenso che gode l'elaborazione teorica presso la comunità scientifica; 4) l'identità, autorevolezza e indipendenza dell'autore della ricerca accreditata e le finalità che persegue, onde valutare se esista una teoria sufficientemente affidabile e in grado di fornire concrete, significative ed attendibili informazioni idonee a sorreggere le argomentazioni probatorie".

<sup>105</sup> QUATTROCOLO, *op. cit.*, 176.

<sup>106</sup> Cfr. HEMPEL, *Philosophy of Natural Science*, 49, 1966; POPPER, *Conjectures and refutations: the Growth of Scientific Knowledge*, 5<sup>a</sup> ed., 1989, 37.

<sup>107</sup> Cfr. STELLA, *op. cit.*, 315 ss.; QUATTROCOLO, *op. cit.*, 147.

<sup>108</sup> KUHN, *The structure of scientific revolution*, Chicago 192, trad. it., *La struttura delle rivoluzioni scientifiche*, Torino 1962.

ta un controllo molto maggiore sugli esperti che non sui normali testimoni»<sup>109</sup>; il giudice, utilizzando i tre metodi di controllo indicati, dovrebbe agire come il rasoio di Occam rispetto alla scienza spazzatura<sup>110</sup>. In tale direzione la più recente giurisprudenza della Suprema Corte richiama l'esigenza di verificare attraverso il contraddittorio, nell'ambito della dialettica processuale «l'attendibilità del perito, l'affidabilità del metodo scientifico utilizzato, e la sua corretta applicazione alla concreta fattispecie processuale (in terminis, sentenze Cozzini e Cantore), operazioni tutte che consentono anche di distinguere le irrilevanti o false opinioni del perito (c.d. *junk science*) dai pareri motivati sulla base di leggi e metodiche scientificamente sperimentate ed accreditate dalla comunità scientifica»<sup>111</sup>.

Nella verifica, in particolare, dell'affidabilità del *software* in base ai criteri esaminati si consideri che, per garantire la parità delle armi tra accusa e difesa, il *software*, come sopra evidenziato, non può essere coperto dal segreto e deve essere stato sottoposto a specifici processi di convalida, consentendo una revisione *ex post* affidabile (che implica che la ricerca empirica sull'argomento specifico sia stata fatta e ripetuta in diverse circostanze, al fine di essere convalidata da colleghi indipendenti); è, inoltre, preferibile che il *software* non sia basato su meccanismi di autoapprendimento (*machine learning solutions*), «che riducono le possibilità di invertire utilmente il processo, al fine di valutarne la logica e l'affidabilità»<sup>112</sup>.

In materia, quindi, sarà importante la convergenza di competenze interdisciplinari tra diagnostica psicologica e criminologica, scienze attuariali e tecnologia, per costruire uno strumento valido di risk assessment, in grado di bilanciare razionalmente i risultati dell'indagine personale con un campione di dati sufficientemente ampio<sup>113</sup>. Si consideri che l'algoritmo può rappresentare, adottando il paradigma multidisciplinare, lo strumento per mettere in comu-

<sup>109</sup> *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993). Tale metodo è stato seguito dalla giurisprudenza seguente, significativa la sentenza della Corte Suprema *General Electric Co. v. Robert K. Joiner*, 78 F.3d 524 (1997).

<sup>110</sup> STELLA, *op. cit.*, 95 s.

<sup>111</sup> Cass., Sez. un., 2 aprile 2019, Pavan, cit.; con commento di BONZANO, *Le Sezioni Unite Pavan e la morte di un dogma: il contraddittorio per la prova spazza via la neutralità della perizia*, in *Dir. pen. proc.*, 2019, 822 ss.; cfr. CONTI, *Scienza controversa e processo penale: la Cassazione e il "discorso sul metodo"*, in *Dir. pen. proc.*, 2019, 848 ss., 860 ss.; MANES, *op. cit.*, 692.

<sup>112</sup> QUATTROCOLO, *op. cit.*, 177. In tale direzione cfr. CONTISSA, LASAGNI, SARTOR, *op. cit.*, 633 che propongono un sistema di certificazione sul modello di quello adottato dal Regolamento 2017/745 (77), che stabilisce i requisiti necessari per ottenere il marchio di Conformità Europea (CE), che certifica un dispositivo medico come prodotto conforme ai requisiti di sicurezza e prestazioni.

<sup>113</sup> D'AGOSTINO, *op. cit.*, 368 ss.

nicazione - riducendo il gap cognitivo -scienze diverse che normalmente non comunicano come «*computational models, artificial intelligence and criminal justice*»<sup>114</sup>.

6. *Il diritto di “accesso” delle parti interessate all’algoritmo nella decisione del Consiglio di Stato, n. 2270/2019.* In materia appare molto interessante una sentenza del Consiglio di Stato che ha affrontato la questione del diritto di “accesso” delle parti interessate all’algoritmo e che, anche se relativa all’utilizzo dell’algoritmo in sede amministrativa, esprime una serie di principi validi anche, e a maggior ragione in considerazione dei beni in gioco, nel contesto penale<sup>115</sup>.

La Sesta Sezione del Consiglio di Stato, in riforma della pronuncia di primo grado, ha accolto il ricorso esperito da un gruppo di docenti avverso una procedura del MIUR di assegnazione dei posti, gestita interamente da un sistema informatico per mezzo di un algoritmo che aveva condotto a provvedimenti irragionevoli, in quanto emanati senza tener conto delle preferenze indicate dalle parti, della graduatoria basata sui meriti, nonché privi di motivazione e in difetto di trasparenza. Le parti, in particolare e in maniera simile a quanto avvenuto nel caso Loomis, ignare delle concrete modalità di funzionamento dell’algoritmo, contestavano la mancanza di trasparenza del procedimento oltre che l’assenza dell’individuazione di un funzionario deputato a valutare specificatamente le singole situazioni e a esternare correttamente le relative determinazioni provvedimentali.

Il Consiglio di Stato, pur riconoscendo l’importanza dell’informatizzazione della Pubblica amministrazione, ha sancito che “in primo luogo, come già messo in luce dalla dottrina più autorevole, il meccanismo attraverso il quale si concretizza la decisione robotizzata (ovvero l’algoritmo) deve essere “*conoscibile*”, secondo una declinazione rafforzata del principio di trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico. Tale conoscibilità dell’algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti. Ciò al fine di poter verificare che gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla

---

<sup>114</sup> QUATTROCOLO, *op. cit.*, 177.

<sup>115</sup> Consiglio di Stato, Sez. VI, sentenza 8 aprile 2019, n. 2270.

stessa amministrazione a monte di tale procedimento e affinché siano chiare - e conseguentemente sindacabili - le modalità e le regole in base alle quali esso è stato impostato”. In altri termini, la “caratterizzazione multidisciplinare” dell’algoritmo (costruzione che certo non richiede solo competenze giuridiche, ma tecniche, informatiche, statistiche, amministrative) non esime dalla necessità che la “*formula tecnica*”, che di fatto rappresenta l’algoritmo, *sia corredata da spiegazioni che la traducano nella “regola giuridica” ad essa sottesa e che la rendano leggibile e comprensibile, sia per i cittadini che per il giudice*” (§ 8.3) (corsivo aggiunto). La Corte non solo afferma il principio della conoscibilità dell’algoritmo ma anche della necessità che sia tradotto in una *regola giuridica conoscibile e comprensibile in piena conformità con il principio di legalità/precisione dell’ordinamento penale*.

Non solo, il Consiglio di Stato pretende ancora che (8.4) la regola algoritmica debba “essere non solo conoscibile in sé, ma anche soggetta alla piena cognizione, e al pieno sindacato, del giudice amministrativo. La suddetta esigenza risponde infatti all’irrinunciabile necessità di *poter sindacare come il potere sia stato concretamente esercitato*, ponendosi in ultima analisi come *declinazione diretta del diritto di difesa del cittadino*, al quale non può essere precluso di conoscere le modalità (anche se automatizzate) con le quali è stata in concreto assunta una decisione destinata a ripercuotersi sulla sua sfera giuridica. Solo in questo modo è possibile svolgere, anche in sede giurisdizionale, *una valutazione piena della legittimità della decisione*; valutazione che, anche se si è al cospetto di una scelta assunta attraverso una procedura informatica, *non può che essere effettiva e di portata analoga a quella che il giudice esercita sull’esercizio del potere con modalità tradizionali*” (corsivo aggiunto). Tanto più chiaramente tale regola circa la necessità che “la regola algoritmica” sia “non solo conoscibile in sé, ma anche soggetta alla piena cognizione, e al pieno sindacato, del giudice” si dovrebbe applicare laddove il giudice penale si avvallesse dell’utilizzo di un algoritmo per valutare la pericolosità sociale in sede penale, anche al fine di garantire - come correttamente evidenziato nella sentenza - il diritto di difesa del cittadino.

Il Consiglio di Stato precisa ancora che “la decisione amministrativa automatizzata impone al giudice di valutare in primo luogo la correttezza del processo informatico in tutte le sue componenti: dalla sua costruzione, all’inserimento dei dati, alla loro validità, alla loro gestione. Da qui, come si è detto, si conferma la necessità di assicurare che quel processo, *a livello amministrativo, avvenga in maniera trasparente, attraverso la conoscibilità dei dati immessi e dell’algoritmo medesimo*. In secondo luogo, conseguente al

primo, il giudice deve poter sindacare la stessa *logicità e ragionevolezza* della decisione amministrativa robotizzata, ovvero della “regola” che governa l’algoritmo, di cui si è ampiamente detto”. Nel caso di specie in base ai principi affermati il Consiglio di Stato ritiene che l’appello debba trovare accoglimento, poiché sussiste “la violazione dei principi di imparzialità, pubblicità e trasparenza, poiché non è dato comprendere per quale ragione le legittime aspettative di soggetti collocati in una determinata posizione in graduatoria siano andate deluse”. La Corte contesta la logicità e razionalità delle scelte realizzate con l’algoritmo; non si tratta, allora, di verificare solo la correttezza dei dati immessi, la trasparenza, ma anche il procedimento logico di funzionamento, le modalità e le regole in base alle quali esso è stato impostato, e quindi la ragionevolezza e correttezza dell’output.

Il Consiglio di Stato non afferma in sé, come aveva fatto la sentenza n. 9227/2018 del Tar Lazio, la necessità di un apporto “umano” al procedimento amministrativo, ma piuttosto, pur riconoscendo la validità e importanza dell’informatizzazione della PA in termini di maggiore efficienza ed economicità, individua le cautele o, meglio, i principi da garantire per conciliare l’utilizzo di algoritmi con il rispetto dei principi, oltretutto di imparzialità e buon andamento della p.a., del *giusto processo*<sup>116</sup>.

7. *Il giudizio individualizzato e il controllo del giudice sull’output.* Per superare i rischi sopra evidenziati di discriminazioni e di generalizzazioni, in cui verrebbero meno le esigenze di un corretto giudizio individualizzato, imprescindibile nell’accertamento della pericolosità sociale in materia penale, si dovrebbe comunque prendere in considerazione solo l’utilizzo di un sistema bifasico, in cui accanto al dataset relativo a profili di criminali ritenuti “simili” a quello dell’imputato, si tiene conto anche delle risposte fornite da questi nel corso dell’interrogatorio (mentre questo non avviene, ad esempio con il PSA, *pretrial risk assessment*)<sup>117</sup>, o comunque di una serie di elementi fattuali relativi al caso e al soggetto concreto; «si tratta di valorizzare il momento “indivi-

<sup>116</sup> GIURDANELLA, *Atti e procedimenti amministrativi informatici: promossa la P.A. Robot, se l’algoritmo è conoscibile*, 29 aprile 2019, in [www.giurdanella.it](http://www.giurdanella.it); in materia CANALINI, *L’algoritmo come “atto amministrativo informatico” e il sindacato del giudice* in *Giornale di diritto amministrativo*, 6, 2019, 781-787; Fasano, *Le decisioni automatizzate nella pubblica amministrazione: tra esigenze di semplificazione e trasparenza algoritmica*, in *Riv. di dir. dei media*, 3, 2019, 234; NICOTRA, VARONE, *L’algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4, 2019, 86 ss.; Strinati, *Algoritmi e decisioni amministrative*, in *Il Foro Amministrativo*, 7-8/2020, pp. 1591-1602.

<sup>117</sup> Cfr. QUATTROCOLO, *op. cit.*, 156.

dualizzante” per frugare i dubbi di una possibile violazione del *right to an individualized sentence*»<sup>118</sup>.

Già dalla vicenda Loomis, inoltre, emerge che il giudice non avrebbe potuto fondare la propria decisione esclusivamente su COMPAS, per «la necessità che l’organo giudicante applichi i risultati del programma facendo esercizio della propria discrezionalità sulla base del bilanciamento con altri fattori». In quel caso il giudice statunitense ritiene che non sussiste alcuna violazione del giusto processo proprio perché la valutazione compiuta utilizzando il sistema di IA costituisce solo un elemento del giudizio, non determinante e corroborato da altri fattori<sup>119</sup>, anche perché “la natura statistico – comparativa del software impiegato non era in grado di garantire un giudizio del tutto personalizzato, ma solo la risultante dell’analisi di circostanze ed esperienze statisticamente simili”<sup>120</sup>.

Per rendere la decisione adottata alla luce dei risultati di un algoritmo conforme ai principi del giusto processo, è sempre necessario il controllo umano sulla decisione algoritmica<sup>121</sup>, come affermato anche nel documento elaborato dal Gruppo di esperti nominato dalla Commissione Europea<sup>122</sup>. La sorveglianza umana è un presidio necessario per garantire l’affidabilità dell’IA e dovrebbe consentire di superare alcuni limiti che incontra la trasparenza del funzionamento dell’algoritmo (come la segretezza tutelata dal diritto di proprietà intellettuale<sup>123</sup> che però nel settore della giustizia dovrebbe venire meno, ma anche la necessità di non rendere il sistema fallace e potenziale oggetto di tentativi di manipolazione o hakeraggio), tenendo conto che “il grado di esplicabilità necessario dipende in larga misura dal contesto e dalla gravità delle conseguenze nel caso in cui il risultato sia errato o comunque impreciso”; “*ad esempio, se le raccomandazioni di acquisto generate da un sistema di IA so-*

<sup>118</sup> D’AGOSTINO, *op. cit.*, 368 ss.

<sup>119</sup> “Nel caso specifico, la lettura delle imputazioni più gravi veniva considerata come elemento di riscontro sufficiente a corroborare la valutazione di rischio generata da COMPAS. Il ricorso veniva quindi rigettato dalla Corte suprema degli Stati Uniti, diventando definitivo nel 2017”, così CONTISSA, LASAGNI, SARTOR, *op. cit.*, 623.

<sup>120</sup> Così OCCHIUZZI, *op. cit.*, 397, infatti nella sentenza si precisa «la valutazione del rischio recidiva da parte di COMPAS non esprime la probabilità specifica che un singolo autore possa nuovamente commettere il reato. Al contrario, fornisce una previsione basata sul confronto tra una serie di dati sul soggetto ed un insieme di informazioni simili» (2 State v. Loomis, 881 N.W.2d par. 15).

<sup>121</sup> CITRON, PASQUALE, *The Scored society*, cit., 6, con riferimento al processo di *algorithmic scoring*; cfr. sul controllo umano High Level Expert Group, *Orientamenti etici per una IA affidabile*, 8 aprile 2019, 18, [ethicsguidelinesfortrustworthyai-itpdf-1.pdf](https://ethicsguidelinesfortrustworthyai-itpdf-1.pdf) (camcom.it).

<sup>122</sup> High Level Expert Group, *Orientamenti etici per una IA affidabile*, cit.

<sup>123</sup> È l’argomento valorizzato nel noto caso Loomis, *Wisconsin Supreme Court, State v. Loomis*, case 2015AP17CR, Judgment July 13th 2016, in *130 Harvard L.R.*, 2017, 1530 ss.

*no imprecise ciò non suscita grandi preoccupazioni etiche, mentre diversa è la situazione quando i sistemi di IA devono valutare l'opportunità o meno di accordare a una persona condannata a pena detentiva la libertà condizionata<sup>124</sup>.*

In tale direzione l'art. 15 della Direttiva 95/46/CE, confluito nell'art. 22 del Regolamento 2016/679/UE, ha stabilito che ogni persona ha il diritto di non essere sottoposta ad una decisione che produca effetti giuridici o abbia conseguenze significative nei suoi confronti, fondata esclusivamente su un trattamento automatizzato di dati destinati a valutare taluni aspetti della sua personalità. La Direttiva 2016/680/UE, poi, all'art. 11, riproduce i contenuti dell'art. 22 del GDPR. L'Italia ha dato attuazione alla predetta Direttiva con il Decreto legislativo 18 maggio 2018 n. 51, il cui art. 8<sup>125</sup> vieta in linea di massima decisioni supportate unicamente da un trattamento automatizzato, compresa la profilazione, che producono effetti negativi nei confronti dell'interessato, «salvo che siano autorizzate dal diritto dell'Unione Europea o da specifiche disposizioni di legge» che, a loro volta, «devono prevedere garanzie adeguate per i diritti e le libertà dell'interessato». In ogni caso, si garantisce, sempre all'art. 8, «il diritto di ottenere l'intervento umano da parte del titolare del trattamento».

In materia penale, in conformità all'art. 111, comma 6, Cost. lo strumento di controllo critico dell'esercizio della discrezionalità vincolata è la motivazione, al centro della quale stanno la ragionevolezza e la fondatezza delle argomentazioni secondo un itinerario di razionalità che deve essere sempre sorretto da un "fondamento ermeneutico controllabile"<sup>126</sup>. L'obbligo di motivazione, giovi ricordarlo, rappresenta una garanzia ordinamentale, un imperativo rivolto ai giudici, che assolve a una duplice funzione di controllo sulla decisione dell'organo giudicante, sia endoprocessuale sia extraprocessuale; il fondamen-

<sup>124</sup> High Level Expert Group, *Orientamenti etici per una IA affidabile*, cit., 15, nota 33.

<sup>125</sup> Art. 8, Processo decisionale automatizzato relativo alle persone fisiche 1. Sono vietate le decisioni basate unicamente su un trattamento automatizzato, compresa la profilazione, che producono effetti negativi nei confronti dell'interessato, salvo che siano autorizzate dal diritto dell'Unione Europea o da specifiche disposizioni di legge. 2. Le disposizioni di legge devono prevedere garanzie adeguate per i diritti e le libertà dell'interessato. In ogni caso è garantito il diritto di ottenere l'intervento umano da parte del titolare del trattamento. 3. Le decisioni di cui al comma 1 non possono basarsi sulle categorie particolari di dati personali di cui all'articolo 9 del regolamento UE, salvo che siano in vigore misure adeguate a salvaguardia dei diritti, delle libertà e dei legittimi interessi dell'interessato. 4. Fermo il divieto di cui all'articolo 21 della Carta dei diritti fondamentali dell'Unione Europea, è vietata la profilazione finalizzata alla discriminazione di persone fisiche sulla base di categorie particolari di dati personali di cui all'articolo 9 del regolamento UE.

<sup>126</sup> Corte costituzionale, v. ad es. la sentenza n. 5 del 2004; MANES, *op. cit.*, 2.

tale controllo endoprocessuale si ricollega alla possibilità per le parti di mettere in discussione quanto deciso dal giudice nei vari gradi del giudizio, e si riconnette al diritto alla motivazione<sup>127</sup>.

Nell'ordinamento italiano, però, l'attivazione di poteri istruttori *ex officio*, o anche su istanza di parte, è preclusa dal disposto dell'art. 220, comma 2, del codice di rito, che prevede che «non sono ammesse perizie per stabilire l'abitudine o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche»; considerando che l'utilizzo di un algoritmo corrisponderebbe a una perizia, sarebbe inutilizzabile per violazione di legge *ex art.* 191, comma 1, c.p.p. L'art. 220 c.p.p. indica infatti alcuni degli elementi che il giudice dovrebbe tenere in considerazione ai sensi del secondo comma dell'art. 133 c.p., escludendo che possano compiersi su di essi accertamenti di carattere tecnico-scientifico<sup>128</sup>.

Il progetto Pisapia adottava un approccio diametralmente opposto che dava spazio alla perizia, ma parte autorevole della dottrina giuridica e medico-legale ha sostenuto la scelta del codice Vassali, evidenziando l'inattendibilità scientifica di una verifica sulla personalità dell'imputato e il rischio di violazione del diritto di difesa perché potrebbero essere aggirate le garanzie e gli strumenti tipici per l'acquisizione della prova<sup>129</sup>; l'art. 133 c.p., rimettendo alla piena discrezionalità del giudice la commisurazione della pena, sarebbe sintomatico della volontà del legislatore di sottrarre campo alle possibili evidenze scientifiche sulla personalità del reo<sup>130</sup>.

Ne deriva che in base a tali disposizioni l'eventuale *output* prodotto dall'IA può essere considerato solo come un mero indizio, che va sempre corroborato con altri elementi di prova e che l'eventuale impiego dell'algoritmo in sede giurisdizionale debba essere assoggettato a un controllo umano significativo<sup>131</sup>.

Fermo restando, però, che lo stesso art. 220, c. 2 fa salvo “quanto previsto ai fini dell'esecuzione della pena o della misura di sicurezza”, aprendo all'utilizzo della perizia e quindi dell'algoritmo, se l'utilizzo dell'algoritmo viene parificato alla perizia, nonché “della psicologia, della criminologia, e delle discipline affini” nella decisione in materia di misure di sicurezza e in sede esecutiva per la scelta delle misure alternative o del tipo di trattamento più

<sup>127</sup> LORUSSO, *Il diritto alla motivazione*, in *Dir. pen. cont.*, 2018, 1 - 2.

<sup>128</sup> D'AGOSTINO, *op. cit.*, 366.

<sup>129</sup> RAMAJOLI, *La prova nel processo penale*, Padova, 1995, 159; D'AGOSTINO, *op. cit.*, 367.

<sup>130</sup> PERCHINUNNO, *Le prove*, in Pisani, Mario et al., *Manuale di procedura penale*, Bologna, 2008, 224; GIANNITTI, *La valutazione della prova penale*, Torino, 2005, 199

<sup>131</sup> QUATTROCOLO, *op. cit.*, 156.

adatto ai particolari bisogni del condannato sulla base dell' "osservazione scientifica della personalità" (art. 1 e 13, l. n. 354/75 e art. 28 d.P.R. n. 431/76) (e aprendo, secondo un certo approccio, al processo bifasico)<sup>132</sup>.

Il tutto assicurando il rispetto delle garanzie del processo penale, a partire, come accennato, dal diritto al contraddittorio<sup>133</sup> (la difesa deve essere messa in condizione di conoscere il peso attribuito ai diversi fattori e di verificare la ragionevolezza dei parametri e del metodo utilizzati, nonché la correttezza della valutazione finale) e il controllo del giudice in base al principio del libero convincimento<sup>134</sup>, giudice che dovrebbe motivare e dare conto della scelta dei dati immessi nel sistema alla luce del fatto concreto.

Tutto ciò pur nella consapevolezza che diversi studi evidenziano il rischio che i giudici si adagino totalmente sul dato fornito dall'algoritmo, limitandosi a validarne le risultanze, anche perché quando gli algoritmi rimpiazzano il processo deliberativo umano, alla decisione finale viene conferita una sorta di "patina di inevitabilità" e, di conseguenza, "di correttezza"<sup>135</sup>; l'algoritmo per la sua presunta oggettività viene equiparato ad una sorta di prova scientifica, rispetto alla quale diventa complicato per il giudice motivare lo scostamento della sua decisione dai risultati della prova, risultando più semplice convalidare tout court gli outputs dell'algoritmo, e scaricare la responsabilità su di esso nel caso in cui la prognosi si riveli sbagliata<sup>136</sup>. La particolare persuasività dell'algoritmo si connette anche al fenomeno noto «nella psicologia cognitiva, come *anchoring*, ossia quella tendenza comune per cui gli individui tendono ad affidarsi al mezzo di prova disponibile, senza avere riguardo alla debolezza esplicativa dello stesso, quando si trovino a dover prendere determinate decisioni»<sup>137</sup>. Il rischio lampante è che il giudice finisca pigramente per accettare il dato emergente dall'algoritmo «concentrandosi cioè più sulla tipologia d'autore che sul fatto oggetto del processo» in una logica propria del diritto penale del nemico, senza prendere in considerazione gli indizi e le risultanze probatorie contrastanti, considerando anche il labile confine tra il principio

<sup>132</sup> LIANI GIARDA, *Art. 220, Oggetto della perizia*, in *Codice di procedura penale commentato*, Art. 1-315, a cura di Giarda, Spangher, V ed., 2017, 2203 e dottrina ivi citata.

<sup>133</sup> MALDONATO, *op. cit.*, 407 ss. sulla scelta degli elementi archiviati, sui loro raggruppamenti e sulle correlazioni dei dati elaborati dall'apparato di intelligenza artificiale, particolarmente in relazione all'oggetto della controversia.

<sup>134</sup> Cfr. sull'importanza del ruolo del giudice in materia MANES, *L'oracolo algoritmico*, cit., 696 ss.; CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Dir. pubbl. comparato ed europeo*, 2019, 124.

<sup>135</sup> CHANDER, *op. cit.*, 1034.

<sup>136</sup> BURCHARD, *op. cit.*, 1930.

<sup>137</sup> MALDONATO, *op. cit.*, 410.

del libero convincimento del giudice e quello del libero arbitrio. Il rischio è anche la compromissione del principio di indipendenza e imparzialità del giudice<sup>138</sup>; si teme il sorgere di una *responsability gap*, in cui le macchine e le loro decisioni automatizzate fungano da comodo capro espiatorio<sup>139</sup>. Si teme, insomma, che i tradizionali rimedi costituiti dalla formazione della prova in contraddittorio e dall'obbligo di motivazione, da soli, non siano in grado di ostacolare la portata espansiva dei risultati dell'algoritmo<sup>140</sup>: il necessario intervento del decisore umano viene considerato «del tutto inefficace di fronte a quella che potremmo definire la travolgente forza pratica dell'algoritmo»<sup>141</sup>.

D'altro canto, però, gli studi di settore dimostrano come la valutazione attuariale del rischio di recidiva del reo sia molto più precisa di quella umana, in quanto riesce a processare una immensa quantità di dati di cui nessun giudice potrebbe ragionevolmente disporre<sup>142</sup>. E, allora, si tratta di valutare se l'utilizzo di simili algoritmi possa alla fine risultare utile a garantire una maggiore attendibilità dei giudizi di pericolosità sociale o se i rischi prevalgano sui possibili benefici, ricordando che «il fulcro di qualsiasi *software* di valutazione del rischio è la teoria psico-criminologica che lo ispira [...] spetta ai progettisti del *software* concepire il modello migliore per fornire una risposta rapida e affidabile in merito alla pericolosità, e soprattutto che nella sua essenza la valutazione digitale del rischio può essere paragonata alla testimonianza di esperti: il *software* però consente un maggior numero di interrogatori, in tempi molto più rapidi, senza bisogno di affidarsi alla presenza personale di un esperto»; e allora «la domanda fondamentale è, quindi, se il *software* di valutazione del rischio debba essere trattato come una perizia psico-criminologica»<sup>143</sup>.

8. *Le prospettive relative all'uso dell'algoritmo in bonam partem*. Considerando, allora, che l'ordinamento penale può trarre un vantaggio dal contributo fornito da altre scienze e dall'utilizzo di algoritmi, ma, allo stesso tempo,

<sup>138</sup> QUATTROCOLO, *op. cit.*, 211.

<sup>139</sup> RODOTÀ, *Il mondo nella rete. Quali i diritti, quali i vincoli*, Laterza, Roma-Bari, VI, 2019, 37; FIORIGLIO, *La "dittatura dell'algoritmo"*, cit., 136.

<sup>140</sup> MALDONATO, *op. cit.*, 407.

<sup>141</sup> «Sia perché sarebbe elevato il rischio di rendere una *probatio diabolica* la dimostrazione dell'unicità di esso come base della pronuncia sia per il già segnalato «mito tecnologico» che potrebbe indurre il giudice a non discostarsi dal risultato della macchina».

<sup>142</sup> BURCHARD, *op. cit.*, 1933 osserva che certi «dubbi sull'uso discriminatorio riguardano solo l'attuazione concreta, ma non le concezioni normativo-ordinamentali fondamentali di una tutela algoritmica efficace ed efficiente dei beni giuridici così come di un'applicazione obiettiva, neutrale e coerente del diritto penale».

<sup>143</sup> QUATTROCOLO, *op. cit.*, 171 s.

deve preservare la sua struttura costituzionale, si ritiene che occorre verificare con molta cautela e nel rispetto del principio di proporzionalità se l'utilizzo di sistemi di Intelligenza Artificiale possa rispondere non solo all'esigenza di incrementare la produttività e l'efficienza del sistema giudiziario, e in particolare l'attendibilità delle valutazioni di pericolosità sociale, ma anche la tutela dei diritti dei cittadini. Ciò richiederebbe una corretta e attenta disciplina per garantire che l'utilizzo della tecnologia rimanga al servizio dell'umanità<sup>144</sup>.

In tale direzione, ad esempio, si dovrebbero stabilire dei limiti all'eventuale utilizzo di algoritmi nel senso che, fermo restando che allo stato il suo utilizzo nell'ordinamento italiano in sede penale potrebbe essere possibile solo nei termini anzidetti in sede di valutazione della pericolosità, semmai de iure condendo se ne dovesse ammettere l'utilizzo anche in sede di commisurazione della pena, se ne dovrebbe limitare l'utilizzo solo in bonam partem. La pena, infatti, seppure deve essere concepita dal legislatore e commisurata dal giudice (nella scelta del tipo e della misura) per perseguire la funzione rieducativa ex art. 27, c. 3 Cost., nondimeno nella sua concreta commisurazione trova il suo limite nel grado di colpevolezza espressa nel fatto commesso. Il rispetto del principio di colpevolezza ex art. 27 Cost. consente l'intervento di esigenze di prevenzione speciale solo al fine di un'eventuale mitigazione dell'intervento punitivo; non sarà possibile aggravare la pena in considerazione di un'accentuata capacità a delinquere ex art. 133, c. 2<sup>145</sup>, e, quindi, del rischio di recidiva, come invece vorrebbe la New Penology nel trattare i condannati come soggetti pericolosi da controllare.

L'eventuale utilizzo di algoritmi, allora, potrebbe essere limitato solo in bonam partem per consentire eventualmente la mitigazione della sentenza in presenza di una limitata capacità a delinquere, nel rispetto del principio della responsabilità penale personale ex art. 27 Cost.; "risk assessment results are used to identify treatment needs or to mitigate sentences, but not to enhance sentences"<sup>146</sup>

L'utilizzo di algoritmi, poi, potrebbe anche servire a identificare il *trattamento* più adatto in considerazione delle caratteristiche del soggetto. La dottrina, infatti, auspica già ora l'utilizzo di sistemi attuariali di valutazione del rischio, come Statist 99R, Stable 2007 e Acute, per la scelta del trattamento più adatto

---

<sup>144</sup> TOMÁS RIZZI, PERA, *Balancing Tests As A Tool To Regulate Artificial Intelligence In The Field Of Criminal Law*, in Special Collection and Artificial Intelligence, 2021, [www.unicr.it](http://www.unicr.it).

<sup>145</sup> Per tutti FIANDACA, MUSCO, *Diritto penale - Parte generale*, Bologna 2018, 804.

<sup>146</sup> ETIENNE, *Legal and Practical Implications of Evidence-Based Sentencing by Judges*, 1 *Chapman J. Crim. Just.* 43, 2009, 50 s.

al livello di rischio e che sia realmente adatto ai bisogni criminogeni dei partecipanti, alla luce di una valutazione condotta secondo un modello evidence-based, come quello cognitivo comportamentale (Risk-Need-Responsivity)<sup>147</sup>. Per dare concretezza alla finalità rieducativa della pena e nell'ottica di superare anche gli effetti discriminatori delle prognosi legate sempre al comportamento passato del soggetto (nella logica del diritto penale del nemico o, della c.d. New Penology come esaminato), - siano esse di carattere intuitivo siano esse compiute avvalendosi di algoritmi predittivi -, si potrebbe, allora, concepire un diverso approccio nella gestione del rischio e, quindi, della pericolosità, che non sia il mero controllo basato su misure detentive o di sorveglianza<sup>148</sup>. Nella logica del principio di proporzionalità e extrema ratio dell'intervento penale, si tratta di adottare un approccio solidaristico proprio di uno Stato sociale, che ha l'obbligo di dare attuazione al principio di uguaglianza sostanziale ex art. 3 Cost., come il c.d. *Supportive Response to Risk*, cioè l'offerta di strumenti di supporto per eliminare i fattori di rischio. Tale prospettiva è già perseguita dai sistemi attuariale di valutazione del rischio, come Statist 99R, Stable 2007 e Acute, per la scelta del trattamento più adatto ai bisogni criminogeni dei partecipanti; da ultimo, il Good Lives Model (GLM)<sup>149</sup> assume un approccio ancora più "supportive" "a strengths-based rehabilitation theory that aims to equip clients with internal and external resources to live a good or better life—a life that is socially acceptable and personally meaningful"<sup>150</sup>.

Un simile modello è stato particolarmente sperimentato nel trattamento della pericolosità di *sex offenders*<sup>151</sup>, purchè il soggetto sia disponibile al cambiamento. Questo modello riparte in generale dal riconoscimento che il "risk is the product of social conditions should lead us to seek responses that directly address those conditions", alla luce di una "socio-structural analysis of risk"<sup>152</sup>,

<sup>147</sup> XELLA, *op. cit.*, 1898 ss.

<sup>148</sup> Le quali, ad esempio nei confronti degli autori di reati sessuali, -rischiano di avere pesanti effetti stigmatizzanti e desocializzanti. Cfr. anche PRESCOTT, ROCKOFF, *Do Sex Offender Registries Make Us Less Safe?*, 2012, [https:// repository.law.umich.edu/articles/83](https://repository.law.umich.edu/articles/83)

<sup>149</sup> XELLA, *op. cit.*, 1898 ss.

<sup>150</sup> MAYSON, *op. cit.*, 2288 ss. Cfr. HANSON, HELMUS, HARRIS, *Assessing the Risk and Needs of Supervised Sexual Offenders: A Prospective Study Using STABLE-2007, Static-99R, and Static-2002R*, in *Criminal Justice and Behavior* 2015.; HELMUS, QUINSEY, FRASER, *Predicting Violent Reoffending with the VRAG-R*, in *The Wiley Handbook of What Works in Violence Risk Management: Theory, Research and Practice*, I ed., a cura di Wormith, Craig, Hogue, 2020

<sup>151</sup> Cfr. XELLA, *op. cit.*, 1898.

<sup>152</sup> MAYSON, *op. cit.*, 2293: "If algorithms targeted the disadvantaged for support rather than for further disadvantage, their effects in the world would be very different. A supportive response to risk would not only serve to prevent new crimes and arrests; it would dramatically mitigate the harm of racial disparity

in questa logica anche l'uso di algoritmi predittivi, per la valutazione della pericolosità e la scelta del trattamento migliore, sarà funzionale non a incrementare un modello di Stato di polizia e repressivo, ma un modello di stato sociale di diritto perché indurrà a implementare non misure restrittive e di controllo sociale, ma misure volte al recupero e al reinserimento sociale.

---

in prediction and, over time, help to mitigate the structural inequalities that give rise to racially disparate risk patterns in the first place”.