DIRITTO PENALE CONTEMPORANEO

QUANDO LA GIUSTIZIA PENALE INCONTRA L'INTELLIGENZA ARTIFICIALE: LUCI E OMBRE DEI *RISK ASSESSMENT TOOLS* TRA STATI UNITI ED EUROPA (*)

di Mitja Gialuz

SOMMARIO: 1. Premessa. – 2. L'inarrestabile diffusione dei *risk assessment tools* negli Stati Uniti. – 3. La valorizzazione dei *risk assessment tools* come fattore decisivo nelle riforme del *bail*. – 4. L'esperienza inglese: l'HART. – 5. La cornice garantistica a livello europeo: la Grande Europa tra Carta etica e diritto di accesso al giudice. – 6. (*segue*): l'impegno per un IA affidabile e il divieto di decisioni basate unicamente su un trattamento automatizzato nel *data protection reform package* dell'Unione europea. – 7. Una prospettiva futuribile anche in Italia?

1. Premessa.

Due anni fa, durante un incontro pubblico, venne chiesto al Presidente della Corte suprema degli Stati Uniti, John Roberts, se potesse prevedere il giorno in cui le *smart machines*, guidate da intelligenze artificiali, potranno assistere il giudice nella ricostruzione del fatto o addirittura intervenire nel processo di *decision-making*. La risposta del giudice Roberts è stata più sorprendente della domanda: «*it's a day that's here*» ha detto, «*and it's putting a significant strain on how the judiciary goes about doing things*»[1].

In effetti, ormai l'intelligenza artificiale (d'ora innanzi IA)[2] non appartiene più alla fantascienza: essa è sempre più presente nella nostra quotidianità, dalle macchine a guida automatica, all'uso del *machine learning* nei servizi di implementazione del

(*) Il contributo riproduce e approfondisce il testo della relazione orale presentata al Congresso annuale dell'Associazione Internazionale di Diritto Penale – Gruppo Italiano su "*Nuove tecnologie e giustizia penale. Problemi aperti e future sfide*", Teramo, 22-23 marzo 2019 (atti in corso di pubblicazione). Si ringraziano i professori Nicola Pisani e Antonino Gullo, organizzatori del convegno, per aver consentito di anticiparne la pubblicazione in questa *Rivista*.

[1] A. LIPTAK, *Sent to Prison by a Software Program's Secret Algorithms*, in *New York Times*, 1° maggio 2017, consultabile a questo *link*.

[2] La definizione di intelligenza artificiale come «*a set of scientific methods, theories and techniques whose aim is to reproduce, by a machine, the cognitive abilities of human beings. Current developments seek to have machines perform complex tasks previously carried out by humans*» si trova in EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, p. 69.

sistema sanitario, dai dispositivi finalizzati a individuare le truffe online, fino agli assistenti domotici come Google Home e Alexa[3]. Questa diffusione si deve al «*rapid development of four self-reinforcing trends*: *ever more sophisticated statistical and probabilistic methods; the availability of increasingly large amounts of data; the accessibility of cheap, enormous computational power; and the transformation of ever more places into IT-friendly environments (e.g. domotics, and smart cities)*»[4].

In tale quadro, non c'è da stupirsi che anche nell'ambito del procedimento penale, che è meccanismo di ricostruzione della realtà, si stia diffondendo, sempre più a macchia d'olio, l'impiego di dispositivi basati sull'IA.

Se si esclude la fase relativa alla prevenzione – nella quale sta crescendo l'impiego di *tools* di *predictive policing*[5], basati sempre di più sull'impiego dei *social media*[6] o su *software* informatici di riconoscimento facciale – ci sono due ambiti nei quali gli strumenti di IA si stanno sviluppando in modo tumultuoso.

Il primo è quello strettamente probatorio, delle cd. *automated o digital evidence* di ultima generazione[7]: nella fase delle indagini, si fa un uso sempre più ampio di sistemi basati su prove algoritmiche in senso lato; e questo impiego sarà destinato a crescere notevolmente con la diffusione dell'*Internet of Things*[8]. Un'altra frontiera è rappresentata dalle cd. *machine-evidence* o *e-evidence* che saranno prodotte dalla stessa automobile a guida automatizzata, nel caso di incidenti generati da una cooperazione tra uomo e robot[9].

L'altro terreno d'elezione per gli strumenti basati sull'IA è quello della cosiddetta "giustizia predittiva", intesa in senso lato.

---

[3] La letteratura in argomento è assai ampia: si leggano, per indicazioni basilari, J. KAPLAN, *Intelligenza artificiale. Guida al prossimo futuro*, Roma, 2017; S. HÉNIN, *AI. Intelligenza artificiale tra incubo e sogno*, Milano, 2019, p. 75 s.

[4] C. CATH – S. WACHTER – B. MITTELSTADT – M. TADDEO – L. FLORIDI, *Artificial Intelligence and the 'Good Society': the US, EU, and UK approach*, in *Science and Eng. Ethics*, 2018, p. 505.

[5] Cfr. L. BENNET MOSES – J. CHAN, *Algorithmic prediction in policing: assumptions, evaluation, and accountability*, in *Policing and Society*, 2016, p. 1. Si vedano anche alcune esperienze italiane, dal *Key crime*, adottato dalla Questura di Milano che ha portato a risultati assai significativi (C. MORABITO, *La chiave del crimine*, consultabile a questo *link*; M. SERRA, *Rapinatore seriale catturato grazie al software "Key crime"*, consultabile a questo *link*), oppure al *X-Law*, un software elaborato dalla Questura di Napoli e completato dal Dipartimento di Pubblica Sicurezza del Ministero dell'Interno e usato in diverse realtà del nostro Paese (cfr. questo *link*).

[6] V. M.S. GERBER, *Predicting crime using Twitter and kernel density estimation*, in *Decision Support Systems*, 2014, vol. 61, p. 115.

[7] Sul rapporto tra *artificial intelligence* e prova penale, si veda il lavoro pionieristico di L. LUPÀRIA, *Prova giudiziaria e ragionamento artificiale: alcune possibili chiavi di lettura*, in *Il concetto di prova alla luce dell'intelligenza artificiale*, a cura di J. Sallantin e J.-J. Szczeciniarz, Milano, 2005, p. XIV ss. Cfr. S. QUATTROCOLO, *Equità del processo penale e* automated evidence *alla luce della Convenzione europea dei diritti dell'uomo*, in *Revista Italo-Española de Derecho Procesal*, 2019, p. 2 s.

[8] U. PAGALLO – S. QUATTROCOLO, *The impact of AI on criminal law, and its twofold procedures*, in *Research Hanbook on the Law of Artificial Intelligence*, a cura di W. Barfield e U. Pagallo, Cheltenham-Northampton, 2018, p. 385.

[9] EUROPEAN COMMITTEE ON CRIME PROBLEMS, *Working Group of Experts on Artificial Intelligence and Criminal Law –Working Paper for the meeting of 27 March 2019*, p. 4.

Per un verso, questa si riferisce in senso proprio all'analisi di un cospicuo numero di pronunce giudiziali effettuato tramite tecnologie di IA, al fine di elaborare previsioni quanto più precise e attendibili in ordine al possibile esito di alcune specifiche tipologie di controversia[10].

Per un altro verso, si stanno diffondendo i cosiddetti *risk assessments tools*, ossia degli strumenti computazionali, spesso fondati sull'IA, in grado di calcolare il rischio che un prevenuto si sottragga al processo o commetta dei reati. Si tratta di veri e propri algoritmi «*that use socioeconomic status, family background, neighborhood crime, employment status, and other factors to reach a supposed prediction of an individual's criminal risk, either on a scale from "low" to "high" or with specific percentages*»[11]. Questi strumenti analizzano un numero molto elevato di dati relativi al passato e individuano delle ricorrenze (ossia dei *pattern*)[12], caratterizzate da una base statistica molto più solida di quelle che stanno al fondo dei giudizi umani[13].

Nella relazione mi concentrerò in particolare su questi strumenti per mettere a fuoco tre aspetti.

Il primo è legato all'utilizzo sempre più massiccio dei *risk assessment tools* negli Stati Uniti e, in particolare, al ruolo cruciale che tali meccanismi stanno avendo nell'ambito della riforma di un istituto chiave dell'ordinamento giuridico *de quo*, ossia il *bail*.

Il secondo riguarda il panorama europeo: si farà riferimento ai primi tentativi di importare anche al di qua dell'Atlantico degli strumenti simili, nonché sui presidi posti in questa materia dal diritto dell'Unione europea e da quello del Consiglio d'Europa.

---

[10] L. VIOLA, voce *Giustizia predittiva* in *Enc. Giur. Treccani, Diritto on line (2018)*, a questo link. Con riguardo all'Italia, C. CASTELLI - C. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*, consultabile a questo link. In termini più ampi, A. GARAPON – J. LASSÈGUE, *Justice Digitale. Révolution graphique et rupture anthropologique,* Paris, 2018. In Europa, la Francia è sicuramente il Paese in cui il dibattito attorno alle tematiche della "giustizia predittiva" è più caldo e vivace. A seguito della approvazione della legge sulla "*République numérique*" del 6 ottobre 2016, al fine di garantire maggiore trasparenza circa l'operato dei Tribunali e delle Corti, l'Amministrazione è tenuta a rendere disponibili on-line tutte le decisioni giudiziali rese sul territorio francese, pari a circa 3 milioni di pronunce all'anno. In conseguenza di ciò, si è venuta a creare un'immensa banca dati *open data*, la quale risulta molto appetibile per le *LegalTech* startup francesi. La startup *Predictice*, ad esempio, ha sviluppato un *software* finalizzato a fornire previsioni circa gli esiti processuali delle controversie, utilizzabile da parte degli avvocati come strumento per l'ottimizzazione delle strategie difensive.

[11] In tal senso, ELECTRONIC PRIVACY INFORMATION CENTER, *Algorithms in the Criminal Justice System*, a questo link. Per quel che riguarda la definizione di algoritmi, cfr. EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, p. 69: «*finite sequence of formal rules (logical operations and instructions) making it possible to obtain a result from the initial input of information. This sequence may be part of an automated execution process and draw on models designed through machine learning*».

[12] Cfr. J. KLEINBERG – H. LAKKARAJU – J. LESKOVEC – J. LUDWIG – S. MULLAINATHAN, *Human Decisions and Machine Predictions*, in *Quarterly Journal of Economics*, 2017, p. 237

[13] «*Imagine a situation where the officer has the benefit of a hundred thousand, and more, real previous experiences of custody decisions? […] no one person can have that number of experiences, but a machine can*» (così, UNIVERSITY OF CAMBRIDGE, *Helping police make custody decisions using artificial intelligence*, 26 febbraio 2018, consultabile a questo link).

L'ultimo profilo è legato alle prospettive di utilizzo di *risk assessments tools* in un paese come il nostro, nel quale, da quasi cent'anni, si esclude il ricorso alla scienza per aiutare il giudice nei giudizi di pericolosità dell'imputato.

2. L'inarrestabile diffusione dei *risk assessment tools* negli Stati Uniti.

Negli ultimi anni si è registrata una vera e propria esplosione dell'uso di algoritmi nella giustizia penale americana[14]. Per rendersene conto, basti pensare che, tra il 2012 e il 2015, 20 leggi in ben 14 Stati «*created or regulated the use of risk assessments during the pretrial process*»[15]. Dal canto loro, varie associazioni molto importanti – tra cui l'*American Bar Association*, la *National Association of Counties*, la *Conference of State Court Administrators*, e la *Conference of Chief Justices* – si sono espresse in favore dell'utilizzo di tali strumenti nella fase pre-processuale[16].

Nel contempo, circa la metà delle giurisdizioni statali fanno oramai uso di meccanismi simili nel *sentencing*, almeno con riguardo a una serie di reati[17]. Anche in questo caso, un report del 2007 del *National Center for States Courts*[18] ha incoraggiato «*this movement towards empirically-informed sentencing approaches*»[19].

Nel sistema penale minorile statunitense vi è stato un analogo *trend* espansivo: se nel 1990 soltanto un terzo degli ordinamenti utilizzava questi *tools*, oggi il numero è cresciuto fino a toccare la soglia dell'86%[20].

Persino il *Model Penal Code* dell'*American Law Institute*, revisionato nel 2017, non ha tralasciato di esprimersi sul punto, esortando a utilizzare «*actuarial instruments or processes to identify offenders who present an unusually low risk to public safety*»[21].

In definitiva, i meccanismi di *risk assessment* vengono oggi impiegati in tutte le fasi del processo penale nordamericano, ogni qualvolta debba essere compiuto un giudizio predittivo: dalle valutazioni sul rilascio del *defendant*, alla fase del *sentencing*, al giudizio sull'applicazione del *parole* o di forme di *probation*[22].

Ciò premesso, va messo in luce che nelle giurisdizioni americane esiste una molteplicità eterogenea di algoritmi predittivi[23], i quali tengono in considerazione

---

[14] Per i dovuti riferimenti dottrinali e giurisprudenziali sul tema si veda l'assai recente articolo di A.Z. HUQ, *Racial Equity in Algorithmic Criminal Justice*, in *Duke Law Journal*, 2019, pp. 1043 ss.

[15] Cfr. A. WIDGERY, NATIONAL CONFERENCE OF STATE LEGISLATURES, *Trends in Pretrial Release: State legislation*, Marzo 2015, accessibile a questo *link*, p. 1.

[16] V. B.L. GARRETT – J. MONAHAN, *Judging Risk*, in *California Law Review*, *Forthcoming*, pp. 10-11.

[17] In tal senso, cfr. B.L. GARRETT – J. MONAHAN, *Judging Risk*, cit., p. 11.

[18] Ci si riferisce a ROGER K. WARREN, *Evidence-Based Practice to Reduce Recidivism: Implications For State Judiciaries* (2007), in http://static.nicic.gov/Library/023358.pdf.

[19] V. ancora B.L. GARRETT – J. MONAHAN, *Judging Risk*, cit., p. 12.

[20] V. B.L. GARRETT – J. MONAHAN, *Judging Risk*, cit., p. 12.

[21] Così, MODEL PENAL CODE: SENTENCING, *Proposed Final Draft*, 10 aprile 2017, p. 171.

[22] Cfr. B.L. GARRETT – J. MONAHAN, *Judging Risk*, cit., p. 9.

[23] Si è calcolato, ad esempio, che nel 2015 erano applicati più di 60 diversi *risk assessment tools* soltanto per la fase del *sentencing*: cfr. A.Z. HUQ, *Racial Equity*, cit., p. 1075.

diversi fattori di rischio "statici" o "dinamici". Per fattore di rischio dinamico si intende «*any factors that contribute to recidivism risk that can change over time*»[24] (si pensi all'età, al lavoro, o all'utilizzo di sostanze psicotrope), mentre i fattori di rischio statici sono, per l'appunto, quelli che non possono variare nel corso del tempo (ad es. il genere e l'età del primo arresto).

Una distinzione fondamentale va compiuta tra algoritmi predittivi elaborati direttamente dai governi statali (o con la loro collaborazione) e *tools* implementati da aziende private.

Ad esempio, lo Stato della Virginia rappresenta il primo ordinamento nord-americano che nel 1994 ha ideato un proprio strumento di *risk assessment*, destinato a essere applicato nella fase del *sentencing*[25]. Nella stessa scia si sono poi posti anche diversi altri Stati, tra cui l'Alabama, l'Alaska, l'Arkansas, la California, la Pennsylvania, la Georgia, l'Indiana, il Montana, il Missouri, l'Ohio e la Nord Carolina.

Uno dei più risalenti e popolari *risk assessment tools* commerciali è, invece, il *Level of Service Inventory – Revised* (LSI-R), sviluppato dall'azienda canadese *Multi-Health Systems.* Tale meccanismo si fonda su molteplici fattori statici e dinamici (tra cui i precedenti penali del soggetto e alcune sue caratteristiche delle personalità) ed è utilizzato quale ausilio per il *sentencing* in alcuni Stati, tra cui il Colorado, la California, l'Iowa, l'Oklahoma e quello di Washington[26].

Un altro celebre *tool* di matrice privata è il *Correctional Offender Mangement Profiling for Alternative Sanction* (COMPAS), elaborato dall'azienda Northpointe (ora Equivant). Il *COMPAS* è un algoritmo, che prende in considerazione le risposte date a un questionario di 137 domande, divise in cinque macro-aree: «*criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion*»[27].

Tale strumento è stato oggetto di profonde critiche: in prima battuta un *report* dell'organizzazione ProPublica ha sostenuto che il *software de quo* sarebbe «*biased against blacks*»[28], posto che lo stesso prende in considerazione alcuni fattori dinamici strettamente correlati alla razza. Più in particolare, si è messo in rilievo come lo strumento crei una disparità di trattamento tra persone di colore e non: difatti, è risultato che le prime sono considerate «*future criminals*» in misura pari al doppio rispetto alle seconde[29]. Da un altro punto di vista, siccome il COMPAS si fonda su algoritmo brevettato e segreto, si è fortemente (e giustamente) criticata la poca

---

[24] Così, D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, in *Berkman Klein Center for Internet & Society, Harvard Law School*, 2017, p. 9.

[25] V. D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System,* cit., p. 11.

[26] Sul punto ELECTRONIC PRIVACY INFORMATION CENTER, *Algorithms in the Criminal Justice System*, cit.

[27] Cfr. D. KEHL – P. GUO – S. KESSLER, *Algorithms in the Criminal Justice System,* cit., p. 11.

[28] Ci si riferisce a J. ANGWIN – J. LARSON – S. MATTU – L. KIRCHNER, *Machine Bias*, in *www.propublica.org*, 23 maggio 2016.

[29] Ci si riferisce a J. TASHEA, *Risk-Assessment Algorithms Challenged in Bail, Sentencing and Parole Decisions*, in *www.abajournal.com*, 1 marzo 2017.

trasparenza del *tool*[30]. Infine, si è messa in discussione la reale capacità predittiva del COMPAS[31].

Nel celebre caso *Loomis*[32], la Corte suprema del Wisconsin ha, però, negato che l'impossibilità per il prevenuto di valutare l'attendibilità scientifica del COMPAS, in ragione della sua segretezza, provochi una lesione del diritto al *due process*: la Corte ha infatti ritenuto che l'imputato potesse, sulla base del manuale d'uso dello strumento, confrontare i dati individuali (ossia gli *input*) e le valutazioni di rischio finali (*output*), confutando dunque l'attendibilità[33]. Peraltro, l'utilizzo del COMPAS è stato ritenuto legittimo solo in presenza di determinati fattori controbilancianti: anzitutto, la Corte ha stabilito che «*a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing*»[34]. In secondo luogo, la Corte ha richiesto che, nel *Presentence Investigation Report* ("PSI"), vengano dati al giudice cinque avvisi, tra i quali merita segnalare il secondo, in forza del quale, siccome la valutazione del rischio si basa su dati riferiti a classi di soggetti, il COMPAS è in grado di identificare gruppi di persone ad alto rischio di recidiva e non un singolo individuo ad alto rischio[35]. Ad ogni modo, laddove sussistano tali condizioni, la Corte ritiene che «*consideration of a COMPAS risk assessment at sentencing along with other supporting factors is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualized sentence*»[36].

---

[30] Così, D. KEHL-P. GUO-S. KESSLER, *Algorithms in the Criminal Justice System,* cit., p. 11.

[31] Si veda lo studio, peraltro un po' risalente, di J. SKEEM – J. ENO LOUDEN, *Assessment of Evidence on the Quality of COMPAS*, 2007, consultabile a questo *link*.

[32] In tale fattispecie l'imputato, che era stato coinvolto in una sparatoria, si era dichiarato colpevole in relazione a due dei cinque capi d'accusa (guida di veicolo senza il consenso del proprietario e tentata violazione di un posto di blocco) e la corte locale lo aveva condannato a sei anni di reclusione e a cinque anni di *extended supervision* basando, almeno in parte, la sua decisione su un giudizio predittivo di «*high risk*» fornito dal COMPAS. A seguito del rigetto di una istanza di *post-conviction release*, il difensore proponeva ricorso alla Corte Suprema, lamentando, anzitutto, la violazione del diritto dell'imputato ad essere valutato sulla base di informazioni accurate; in secondo luogo, la violazione del diritto ad una sentenza individualizzata e, infine, l'appartenenza al genere maschile tra i vari dati utilizzati per valutare la pericolosità (per un'accurata ricostruzione, si veda S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale 'predittiva'*, in corso di pubblicazione su *Cass. pen.*). Il caso ha destato grande scalpore nell'opinione pubblica e, nella vulgata generale, è divenuto uno degli esempi paradigmatici di sostituzione della macchina all'uomo: v., in particolare, A. LIPTAK, *Sent to Prison by a Software Program's Secret Algorithms*, in *The New York Times*, 1.5.2017, consultabile a questo *link*.

[33] State v. Loomis, 881 NW 2d 749 (Wis 2016), § 53-54. Per un commento alla sentenza v. *Criminal Law – Sentencing Guidelines – Wisconsin Supreme Court Requires Warnings before Use of Algorithmic Risk Assessment in Sentencing – State v. Loomis*, in *Harward Law Review*, 2017, pp. 1530 ss. L'utilizzo del COMPAS in fase di *sentencing* era già stato ammesso in Wisconsin dalla sentenza State v. Samsa, 2015 WI App 6.

[34] State v. Loomis, cit., § 99. Laddove, infatti, «*a COMPAS risk assessment were the determinative factor considered at sentencing this would raise due process challenges regarding whether a defendant received an individualized sentence*» (§ 68).

[35] State v. Loomis, cit., § 100.

[36] State v. Loomis, cit., § 72.

Dal canto suo, la Corte suprema USA ha confermato tale decisione, rigettando il *writ of certiorari* presentato avverso la stessa[37]. Queste pronunce vanno dunque ad avallare la letteratura secondo cui la creazione di un algoritmo, effettuata in modo appropriato, può andare a perfezionare le decisioni predittive dell'uomo, che sono naturalmente basate su una limitata esperienza: i *tools* contribuirebbero, infatti, a ridurre la popolazione carceraria e ad assicurare l'eliminazione delle disparità razziali, diventando, così, «*a force for racial equity*»[38].

Infine, il PSA (*Public Safety Assessment*) è stato creato dalla Laura and John Arnold Foundation, utilizzando i *reports* di 750.000 casi, riguardanti oltre 300 giurisdizioni americane[39], proprio per assicurare la trasparenza del funzionamento ed eliminare gli effetti discriminatori del COMPAS, escludendo l'incidenza negativa dei dati riguardanti le condizioni economiche, razziali e di genere. A supporto della decisione di rendere pubblico il funzionamento di tale algoritmo, Matt Alsdorf, uno dei rappresentanti della *Arnold Foundation*, ha affermato che «*it's important from a fairness per-spective for all the parties to understand that goes into a risk assessment*»[40].

Il PSA è senza dubbio uno dei *tools* più utilizzati per la fase *pre-trial*, essendo adottato da dozzine di giurisdizioni statunitensi e da tre Stati: l'Arizona, il Kentucky e il New Jersey[41]. Tale meccanismo esamina nove fattori, legati all'età del prevenuto, l'imputazione e i suoi precedenti penali per determinare due fattori di rischio: da un lato, il pericolo che il prevenuto non si presenti in udienza e, da un altro lato, la probabilità che questi commetta un reato se rilasciato prima del dibattimento[42].

Ciascuno dei parametri ha un diverso peso nel funzionamento dell'algoritmo; ad esempio, il numero delle precedenti condanne viene valutato in misura maggiore rispetto agli altri criteri. Peraltro, si è in presenza in questo caso di dati neutrali e, quindi, si potrebbe pensare che ciò elimi il problema della disparità di trattamento. Tuttavia, «*there is still a fear that the PSA and other similar tools perpetuate racial biases and encourage the pretrial release of dangerous criminals*»[43]. In merito, Jeremy Travis, uno dei rappresentanti della Arnold Foundation, ha però precisato non solo che, nel costruire l'algoritmo, si è cercato di essere certi che lo stesso fosse «*race neutral*», ma anche che l'ultima parola spetta al giudice, il quale ben si potrebbe discostare dai risultati dell'algoritmo, qualora non si adattino al caso concreto[44].

Ad ogni modo, il PSA ha fornito dei riscontri positivi: nel *Lucas County*, Ohio, – che ha adottato il *software* nel 2015 – si sono accertati, da un lato, un aumento del

---

[37] 137 S. Ct. 2290.

[38] Cfr. J. KLEIBERG – H. LAKKARAJU – J. LESKOVEC – J. LUDWIG – S. MULLAINATHAN, *Human Decision and Machine Predictions*, in *The Quarterly Journal of Economics*, 2018, p. 237.

[39] Sul punto si veda la descrizione del *tool*, pubblicata in https://www.psapretrial.org/about/factors.

[40] J. TASHEA, *Risk-Assessment Algorithms*, cit.

[41] Si veda la scheda *About the PSA*, pubblicata in https://www.psapretrial.org/about.

[42] Cfr. *About the PSA*, cit.

[43] Si veda K. PATRICK, *Arnold Foundation to Roll Out Pretrial Risk Assessment Tool Nationwide*, in *www.insidesources.com*, 3 settembre 2018.

[44] Così, ancora, K. PATRICK, *Arnold Foundation to Roll Out Pretrial Risk Assessment Tool Nationwide*, cit.

numero di persone messe in libertà, senza ricorrere al *bail*, e, dall'altro, una diminuzione del numero di reati commessi in attesa di giudizio[45].

Ciò nonostante, si continuano a mettere in luce le problematiche insite nell'utilizzo degli algoritmi: si è affermato, infatti, che «*even if an algorithm is equally accurate for all, more blacks and males will be classified as high risk because African-Americans and men are more likely to be arrested for a violent crime*»[46].

3. La valorizzazione dei *risk assessment tools* come fattore decisivo nelle riforme del *bail*.

Come noto, in quasi ogni giurisdizione degli Stati Uniti è utilizzato l'istituto del *bail*[47], quale condizione per rilasciare un soggetto in attesa dello svolgimento del giudizio a suo carico[48]. Siffatto meccanismo è stato però negli ultimi tempi oggetto di numerose e accese critiche provenienti non solo dalla dottrina, ma da buona parte dell'opinione pubblica americana[49]. In particolare si è a più voci sostenuto, che il *bail*: a) provoca tremendi costi non solo ai prevenuti, alle loro famiglie, ma anche alla collettività; b) contribuisce in maniera determinante alla situazione critica di utilizzo spropositato della detenzione preventiva negli USA; c) crea insopportabili disuguaglianze tra persone abbienti e non abbienti[50].

Per rendersi conto di quanto la tematica della riforma del *bail* risulti di stretta attualità, basti pensare che gli *editorial boards* di tre dei più importanti quotidiani statunitensi – il *Los Angeles Times*, il *Washington Post* e il *New York Times* – si sono tutti schierati a favore dell'abolizione dello stesso[51]. Nel contempo, numerose organizzazioni diffuse nel paese si battono ufficialmente per riformare tale meccanismo, tra cui si possono ricordare l'*American Bar Association*, la *National Association of Pretrail Services Agencies*, la *Conference of State Court Administrators*, la *National Association of Counties*, la *Conference of Chief Justies*, l'*American Jail Association*, l'*International Association of Chiefs of Police*, l'*Association of Prosecuting Attorneys* e la *National Association of Criminal Defence Lawyers*[52].

---

[45] V. J. TASHEA, *Risk-Assessment Algorithms*, cit.

[46] In tal senso, ancora, J. TASHEA, *Risk-Assessment Algorithms*, cit.

[47] Per una descrizione di tale meccanismo cfr. V. TONDI, *Il* Bail. *La libertà su cauzione negli ordinamenti anglosassoni*, Padova, 2016.

[48] Si veda C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform. A Guide for State and Local Policymakers*, *Criminal Justice Policy Program*, Harvard Law School, febbraio 2019, p. 1.

[49] *Note. Bail Reform and Risk Assessment: the Cautionary Tale of Federal Sentencing*, in *Harvard Law Review*, 2018, pp. 1125 ss.

[50] *Note. Bail Reform and Risk Assessment*, cit., p. 1125.

[51] Editoriale del 16 agosto 2017, *How the Poor Get Locked Up and the Rich Go Free*, in *L.A. TIMES*, consultabile a questo *link*; Editoriale del 9 settembre 2017, *Fixing the Unfair Bail System Is Worth the Costs*, in *WASH. POST*, consultabile a questo *link*; Editoriale del 25 agosto 2017, *Cash Bail's Lonely Defender*, in *N.Y. TIMES*, consultabile a questo *link*.

[52] Cfr. C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 9.

Anche dal punto di vista politico la necessità di intervenire sul *bail* sta incontrando un supporto trasversale: il Senatore democratico Kamala Harris e il repubblicano Rand Paul hanno di recente introdotto delle norme federali volte a incoraggiare gli Stati a intervenire sul punto e il Senatore Barnie Sanders ha introdotto una «*legislation to eliminate money bail at the federal level*»[53].

Questo ampio movimento sta producendo i suoi frutti: numerosi Stati – tra cui quello di Washington, il Kentucky, il New Jersey, l'Illinois e la California – hanno approvato importanti novelle del *bail*[54].

In questa sede, interessa porre in luce che i legislatori, onde fornire un ausilio ai giudici nella valutazione del rischio che il prevenuto commetta un crimine nel corso della regiudicanda o si sottragga al processo e, nel contempo, ridurre il numero dei detenuti in attesa di giudizio, «*have adopted algorithmic risk assessment tools as part of their pretrial reforms*»[55].

I due Stati che hanno approvato le riforme più rilevanti sul punto sono il Kentucky e la California.

Il Kentucky ha adottato un progetto pilota – denominato "*Administrative Pretrial Release Program*" – in 20 giurisdizioni su 120, poi esteso nel 2017 all'intero Stato, basato su un utilizzo peculiare del PSA. Al fine di incrementare l'efficienza del sistema e salvare le risorse per i prevenuti più pericolosi, in tale Paese si prevede che per una serie di reati i *pretrial officers* possano ordinare il rilascio immediato dei prevenuti, il cui rischio di fuga e commissione di reati risulti sulla base del *tool* in questione basso o moderato, senza l'intervento di un giudice[56]. In altri termini, un prevenuto può ottenere il rilascio immediato, senza *bail* e senza che venga svolta un'udienza da parte di un giudice, se: a) il suo punteggio al PSA è basso e moderato; b) si procede nei suoi confronti per un *misdemeanor*, che non abbia natura violenta o sessuale. Si è anche stabilito che tramite l'approvazione di norme locali è possibile ampliare l'ambito di applicazione di questa procedura semplificata di rilascio per alcuni *felony* (sempre caratterizzati da una natura non violenta e non sessuale).

Ancora più significativa è la riforma californiana: nel 2018 tale Stato è diventato il primo ordinamento statunitense ad abolire del tutto il *bail*, sostituendolo «*with "risk assessments" of individuals and non-monetary conditions of release*»[57].

Tale novella, discussa da tempo e diventata inevitabile dopo che all'inizio del 2018 la *California appellate court* aveva dichiarato incostituzionale il sistema di *cash bail* previamente in vigore[58], è stata salutata con grande entusiasmo dal Governatore Brown, il quale, pochi momenti dopo aver firmato il *California Money Bail Reform Act*,

---

[53] Sul punto cfr. C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 9.
[54] Cfr. C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 9.
[55] V. C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 14.
[56] Sul punto cfr. C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 40.
[57] Cfr. C. KALMBACHER, *California Just Eliminated Cash Bail – Here's What That Means*, consultabile a qeusto link, 28 agosto 2018.
[58] Ci si riferisce a Court Of Appeal Of The State Of California, First Appellate District Division Two, *In re Kenneth Humprey, on Habeas Corpus*, 25 gennaio 2018.

ha affermato: «*today, California reforms its bail system so that rich and poor alike are treated fairly*»[59].

È interessante notare che, in modo analogo a quanto accade in Kentucky, anche in California una norma stabilisce che il *Pretrial Assessment Services* è tenuto a rilasciare *direttamente, senza l'intervento di un giudice*, i prevenuti il cui livello di rischio di fuga risulti basso, a seguito dello svolgimento di un *test* per il tramite di uno dei *risk assessment tools* accreditati, contenuti in una lista stilata dal *Judicial Council* della California[60].

Com'era peraltro prevedibile, la riforma in questione, la cui entrata in vigore era originariamente prevista nell'ottobre del 2019, ha raccolto molte critiche, soprattutto per il suo affidarsi in modo così marcato su uno strumento tanto controverso come i *risk assessment tools*[61]. Di talché, una coalizione di *bail bond industry* ha avuto gioco facile nel raccogliere le firme necessarie per lo svolgimento di un *referendum* abrogativo sulla novella *de qua*, che si terrà nel novembre del 2020, fino allo svolgimento del quale l'entrata in vigore dell'*Act* rimane congelata[62].

4. L'esperienza inglese: l'*HART*

L'impiego dei *tools* di *risk assessment* non è limitato agli Stati Uniti. Vi sono alcune esperienze rilevanti anche in Europa, anche se non si è raggiunta una diffusione paragonabile a quella d'oltreoceano.

La sperimentazione più significativa è probabilmente quella sorta in Inghilterra, dove la polizia del Durham, in collaborazione con l'università di Cambridge, ha messo a punto un sistema denominato *Harm Assessment Risk Tool* (d'ora innanzi HART), con l'obiettivo di promuovere processi decisionali coerenti che permettano di realizzare interventi mirati a ridurre il rischio di recidiva[63]. In particolare, tale strumento è stato utilizzato dal corpo di polizia di Durham a partire dal 2017[64] in chiave di *diversion*, ossia al fine di valutare quando una persona può essere sottoposta a un *rehabilitation programme*, chiamato *Checkpoint*, il quale costituisce un'alternativa all'esercizio dell'azione penale[65].

---

[59] Cfr. A. DOBUZINSKIS, *California scraps cash bail in move touted as economic fairness*, accessibile a questo *link*.
[60] Cfr. Senate Bill No. 10, Chapter 244, Article 4, 1320.10 (b).
[61] V. Human Rights Watch Opposes California Senate Bill 10, The California Bail Reform Act, accessibile a questo *link*.
[62] Cfr. *California Replace Cash Bail with Risk Assessments Referendum (2020)*, accessibile a questo *link*.
[63] Cfr. M. OSWALD – J. GRACE-S. URWIN – G.C. BARNES, *Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality*, in *Information and Communications Technology Law*, 2018, p. 227.
[64] V. BIG BROTHER WATCH, *Big Brother Watch's written evidence in the justice system for the Law Society's system for the Law Policy Commission*, accessibile a questo *link*.
[65] Cfr. M. OSWALD – J. GRACE – S. URWIN – G.C. BARNES, *Algorithmic risk assessment policing models*, cit., p. 227.

Più specificamente, questo modello è un «*machine learning tool*»[66] che svolge dei giudizi predittivi per verificare il rischio che un soggetto arrestato commetta dei reati nei due anni successivi. Sulla base del risultato, la persona viene catalogata come ad alto, moderato o basso rischio, a seconda della previsione per cui essa perpetuerà un grave reato – come ad esempio un omicidio, una violenza o una rapina –, un reato considerato non grave, o infine, nessuna fattispecie delittuosa. Ebbene, il programma *Checkpoint* è destinato esclusivamente alla seconda categoria, vale a dire quella a moderato rischio[67].

L'HART è stato creato sulla base dell'analisi di circa 104.000 casi avvenuti a Durham in un arco temporale di cinque anni, dal 2008 al 2012[68]; esso, inoltre, si fonda su una particolare forma di *machine learning*, chiamata *random forest*[69], che prende in considerazione 34 variabili, 29 delle quali collegate alla storia criminale del soggetto, unitamente all'età, al genere, nonché a due codici postali di residenza[70].

Orbene, proprio con riferimento a quest'ultimo tipo di variabile sono sorte delle rilevanti critiche in termini di *privacy*[71], ma non solo.

Si è rilevato, infatti, che uno dei due codici postali – il *Mosaic code* – è stato elaborato dalla polizia di Durham mediante l'utilizzo della piattaforma informatica *Mosaic*, gestita da una compagnia privata di *marketing*, *Experian*[72]. Più specificamente, tale strumento costituisce un *geodemographic segmentation tool*, il quale profila 50.000.000 di persone in tutto il Regno Unito in 66 categorie. I dati presi in considerazione sono oltre 850.000.000, ed essi possono essere i più disparati: ad esempio, la composizione famigliare, l'occupazione della persona, la salute, i consumi di gas ed elettricità, nonché *online data*[73].

Nella pagina di presentazione del sistema si afferma che «*Mosaic enables consistent targeting across a multitude of on and off-line channels*»[74]; difatti, vi è una parte dedicata proprio alle attività *online* di ciascun gruppo, tra cui informazioni sui siti visitati, nonché sulla frequenza di utilizzo dei vari strumenti tecnologici di comunicazione e dei *social media*[75].

Sulla base di questa preoccupante realtà, si è quindi incisivamente sostenuto che «*this tool raises novel questions about big data and privacy, the right to be free from*

---

[66] Cfr. S. URWIN, *Written evidence submitted by Sheena Urwin, Head of Criminal Justice, Durham Constabulary*, in *www.parliament.uk*, 20 febbraio 2018.

[67] V. M. OSWALD – J. GRACE-S. URWIN – G.C. BARNES, *Algorithmic risk assessment policing models*, cit., p. 227.

[68] Cfr. M. OSWALD – J. GRACE – S. URWIN – G.C. BARNES, *Algorithmic risk assessment policing models*, cit., p. 228.

[69] V. M. OSWALD – J. GRACE – S. URWIN – G.C. BARNES, *Algorithmic risk assessment policing models*, cit., p. 227.

[70] Cfr. H. COUCHMAN, *Policing by Machine. Predictive Policing and the Threat to Our Rights*, in *www.libertyhumanrights.org.uk*, gennaio 2019; M. OSWALD – J. GRACE – S. URWIN – G.C. BARNES, *Algorithmic risk assessment policing models*, cit., p. 228.

[71] V. H. COUCHMAN, *Policing by machine. Predictive policing*, cit.

[72] Cfr. BIG BROTHER WATCH, *Big Brother Watch's written evidence in the justice system*, cit.; ID., *A Closer Look at Experian Big Data and Artificial Intelligence in Durham Police*, accessibile a questo link, 6 aprile 2018; ID., *Police use Experian Marketing Data for AI Custody Decisions*, a questo link, 6 aprile 2018.

[73] Cfr. BIG BROTHER WATCH, *Big Brother Watch's written evidence in the justice system*, cit.

[74] Si veda la pagina di presentazione di *Mosaic* a questo link.

[75] V. BIG BROTHER WATCH, *A Closer Look at Experian Big Data and Artificial Intelligence*, cit.

profiling and automated decisions, algorithmic discrimination, and fairness in the criminal justice system»[76]. Ancora, si è affermato come sia inaccettabile che questo strumento venga utilizzato «*to inform potentially life-changing criminal justice decisions*»[77].

5. La cornice garantistica a livello europeo: la Grande Europa tra Carta etica e diritto di accesso al giudice.

Dinnanzi alla diffusione di tali strumenti, il tema non può e non deve essere se si è a favore o contro di essi. Il dibattito reale, che è stato opportunamente avviato a livello europeo, ha a oggetto il come i sistemi giudiziari saranno in grado, nel prossimo futuro, di far fronte a questi sviluppi tecnologici, senza divenirne vittime, e di inquadrare il loro utilizzo per assicurare il rispetto dei diritti fondamentali.

Sul versante della Grande Europa, vi è un'attenzione straordinaria per il crescente impiego di strumenti digitali anche in sede giudiziaria. Nel marzo del 2018 veniva pubblicato uno studio su *Algorithms and Human Rights*[78], che ha costituito una base importante per l'adozione, nel dicembre dello stesso anno, di un documento di *soft law* particolarmente significativo. Si tratta della Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia, la quale è stata adottata dalla Commissione per l'efficienza della giustizia (CEPEJ)[79]. Il documento è rivolto ai «*public and private stakeholders responsible for the design and deployment of artificial intelligence tools and services that involve the processing of judicial decisions and data*», nonché ai «*public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services*»[80]. Esso fissa cinque principi generali.

Anzitutto, quando gli strumenti di IA vengono impiegati come ausilio nei processi, si deve assicurare che non violino il diritto di accesso al giudice e il diritto a un processo equo (parità di armi e rispetto del contraddittorio).

In secondo luogo, viene sancito il canone di non discriminazione: considerata la capacità di questi metodi di elaborazione di rivelare le discriminazioni esistenti, i soggetti pubblici e privati devono garantire che essi non riproducano o aggravino tali discriminazioni e che non conducano ad analisi deterministiche. Ciò vale, in particolare, quando vengano in rilievo dati sensibili, quali quelli relativi all'origine razziale o etnica, al *background* socio-economico, alle opinioni politiche, alle convinzioni religiose o filosofiche, all'appartenenza sindacale, o ancora i dati genetici, biometrici,

---

[76] Cfr. BIG BROTHER WATCH, *Big Brother Watch's written evidence in the justice system*, cit.

[77] V. BIG BROTHER WATCH, *Big Brother Watch's written evidence in the justice system*, cit.

[78] Il riferimento è ad *Algorithms and Human Rights - Study on the human rights dimension of automated data processing techniques and possible regulatory implications*, disponibile a questo *link*.

[79] Per un commento a prima lettura, cfr. S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *www.lalegislazionepenale.it*, 18 dicembre 2018.

[80] EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, 3-4 dicembre 2018, p. 5.

relativi alla salute o quelli riguardanti la vita sessuale o l'orientamento sessuale. Quando tale discriminazione è stata identificata, occorre prendere in considerazione misure correttive per limitare o, se possibile, neutralizzare questi rischi e anche sensibilizzare le parti interessate.

Il terzo principio affermato dalla Carta riguarda la qualità e sicurezza: per un verso, si raccomanda di utilizzare esclusivamente dati (in particolare decisioni giudiziarie) provenienti da fonti certificate; per altro verso, il processo deve essere tracciabile e i modelli e gli algoritmi creati devono poter essere memorizzati ed eseguiti in ambienti sicuri, in modo da garantire l'integrità del sistema.

Il quarto canone è essenziale ai nostri fini in quanto si prescrive la trasparenza, l'imparzialità e la correttezza: rispetto alle esigenze di tutela della proprietà intellettuale devono prevalere l'accessibilità al processo algoritmico, l'assenza di pregiudizi e l'integrità intellettuale. Questi valori possono essere assicurati anzitutto con la completa trasparenza tecnica (del codice sorgente e della documentazione), la quale peraltro non appare di per sé sufficiente: si è correttamente notato che, «anche là dove il *reverse engeneering* sia possibile, la comprensione del modello rimane questione limitata ai soli esperti, con esclusione degli effettivi destinatari della 'decisione automatizzata'»[81]; è indispensabile ed urgente creare dunque autorità pubbliche indipendenti che possano valutare e certificare i *tools a priori* e poi monitorarne il funzionamento.

L'ultimo canone è quello denominato *under user control*, in forza del quale va escluso un approccio prescrittivo dell'impiego dell'IA e va assicurato che gli utilizzatori agiscano come soggetti informati e che abbiano il pieno controllo delle loro scelte. Premesso che l'utente può essere, sia l'operatore del diritto che utilizza il *tool*, sia l'interessato destinatario della decisione, tale principio si traduce, per il primo, nella possibilità di riesaminare le decisioni e i dati utilizzati per produrre un risultato e continuare a non esserne necessariamente vincolati alla soluzione suggerita dal dispositivo di IA, alla luce delle caratteristiche peculiari del caso specifico. Per l'utente, invece, nel diritto di essere informato delle diverse opzioni disponibili e nel diritto alla consulenza legale e all'accesso a un giudice ai sensi dell'art. 6 C.e.d.u.

Nella prima appendice della Carta – contenente uno Studio sull'impiego dell'IA nei sistemi giudiziali – vengono, per un verso, ribadite le criticità legate a possibili effetti discriminatori degli strumenti predittivi di responsabilità e, per l'altro, si raccomanda il rispetto del principio di parità delle armi, della presunzione di innocenza e si sottolinea la necessità che il soggetto interessato abbia la possibilità di contestare la validità scientifica dell'algoritmo e il peso attribuito ai vari dati: in quest'ottica, la chiave è rappresentata dal diritto di accesso al giudice, che trova un suo fondamento anche nei principi di protezione dei dati personali[82].

---

[81] Con queste parole, S. QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., p. 8.

[82] Cfr. EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter*, cit., Appendix I, *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, § 138.

Nella seconda appendice, proprio in considerazione degli effetti discriminatori e deterministici che hanno avuto i *risk assessment tools* negli Stati Uniti, si inquadrano tali dispositivi nella categoria di quelli dei quali si auspica un utilizzo con le più estreme riserve[83].

Sul versante della Convenzione, non c'è dubbio che, laddove si prospetti un utilizzo dei *risk assessment tools* nella materia della libertà personale, vengono in gioco le garanzie dell'art. 5 C.e.d.u. e, in particolare, il diritto all'accesso al giudice, nelle sue due forme contemplate dal par. 3 e dal par. 4, nonché l'obbligo di motivazione.

Come noto, la prima riguarda il diritto di essere tradotti davanti a un giudice: tradizionalmente, facendo leva sul testo inglese secondo il quale il soggetto arrestato «*shall be brought promptly before a judge*», si è ritenuto che tale norma contempli il diritto a un vero e proprio "contatto fisico" con il giudice[84]: evidentemente, nell'epoca dell'intelligenza artificiale, questa sottolineatura assume un significato nuovo; la decisione sulla libertà personale di un imputato presunto innocente va presa personalmente da un giudice. Non si può pertanto nemmeno ipotizzare una soluzione analoga a quella della California o del Kentucky. Peraltro, va ricordato che ogniqualvolta la Convenzione richiama l'intervento del giudice ciò significa che devono essere assicurate le garanzie essenziali di un procedimento giudiziario[85], con la fissazione di un'udienza alla quale deve partecipare personalmente il detenuto, il quale deve essere a conoscenza dei motivi che giustificano la sua detenzione[86] e deve essere assistito da un difensore[87].

La seconda forma di accesso al giudice è costituita dal diritto a un ricorso effettivo davanti al tribunale per un controllo sulla legittimità della misura. Un procedimento che implica l'applicazione delle garanzie del *fair trial*[88], sia pure all'interno di una procedura che deve concludersi "in breve tempo"[89]. Questo significa, in particolare, che vanno rispettati il canone della parità delle parti[90] e che deve essere consentito alla difesa l'accesso ai documenti investigativi in quanto, in mancanza della conoscenza dei dati che giustificano l'arresto, il diritto di esperire un ricorso per la verifica della legalità della detenzione si riduce in una mera formalità[91]. Con riguardo

---

[83] Cfr. EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter*, cit., Appendix II, *Which uses of AI in European judicial systems?*, p. 52.

[84] Corte e.d.u., 21 dicembre 2000, *Egmez c. Cipro*, § 90; Corte e.d.u., 4 dicembre 1979, *Schiesser c. Svizzera*, § 31.

[85] Corte e.d.u., 26 maggio 1993, *Brannigan e McBride c. Regno Unito*, § 58

[86] Corte e.d.u., 28 ottobre 1998, *Assenov e a. c. Bulgaria*, § 146; Corte e.d.u., 26 giugno 1991, *Letellier c. Francia*, § 35.

[87] Nel senso della necessaria presenza del difensore, Corte e.d.u., 14 ottobre 2010, *Brusco c. Francia*, § 45.

[88] Corte e.d.u., 31 gennaio 2002, *Lanz c. Austria*, § 41; Corte e.d.u., 13 febbraio 2001, *Schöps c. Germania*, § 44.

[89] J. MURDOCH, *L'article 5 de la Convention européenne des droits de l'homme*, Strasburgo, 2004, p. 102.

[90] Corte e.d.u., 20 gennaio 2004, *G.K. c. Polonia*, § 91; Corte e.d.u., 25 giugno 2002, *Migón c. Polonia*, §§ 79-80; Corte e.d.u., 31 gennaio 2002, *Lanz c. Austria*, § 44; Corte e.d.u., 25 giugno 2002, *Migón c. Polonia*, §§ 79-80.

[91] Corte e.d.u., 30 marzo 1989, *Lamy c. Belgio*, § 29; per una ricapitolazione dei principi, Corte e.d.u., GC, 9 luglio 2009, *Mooren c. Germania*, § 108 e ss.

agli strumenti di IA, evidentemente occorrerebbe chiarire alla difesa qual è il funzionamento dell'algoritmo che sta al fondo del *risk assessment tool*.

La terza garanzia fondamentale connessa con l'accesso al giudice è rappresentata dall'obbligo di motivazione: la giurisprudenza consolidata di Strasburgo richiede la motivazione per ogni pronuncia attinente allo *status libertatis*: «*justification for any period of detention, no matter how short, must be convincingly demonstrated by the authorities*»[92].

6. (*segue*): l'impegno per un IA affidabile e il divieto di decisioni basate unicamente su un trattamento automatizzato nel *data protection reform package* dell'Unione europea.

In forza dell'art. 52, par. 3, della Carta di Nizza, le stesse garanzie in materia di libertà personale si estendono anche sul versante dell'Unione europea, che riconosce il diritto alla libertà e alla sicurezza, nell'art. 6 della stessa Carta dei diritti fondamentali[93]. A questi presidi connessi al diritto di accesso al giudice – e ai suoi corollari – si aggiungono due aspetti fondamentali.

Per un verso, si deve segnalare l'impegno della Commissione europea per garantire lo sviluppo di un'intelligenza artificiale affidabile: nel dicembre 2018 sono state pubblicate le "*Draft Ethics Guidelines for Trustworthy AI*", elaborate da un Gruppo di esperti ad alto livello sull'IA e aperte alla consultazione pubblica[94]. Si tratta di un documento che parte dalla consapevolezza che l'IA ha la capacità di generare enormi vantaggi per gli individui e per la società, ma comporta anche determinati rischi che vanno gestiti in modo adeguato: occorre dunque assicurare di seguire la strada che massimizza i benefici dell'IA riducendone al minimo i rischi. In quest'ottica, si afferma la necessità di assicurare che «*AI is human-centric: AI should be developed, deployed and used with an "ethical purpose" (…), grounded in and reflective of fundamental rights, societal values and the ethical principles of Beneficence (do good), Non-Maleficence (do no harm), Autonomy of humans, Justice, and Explicability*»[95]. Un approccio antropocentrico all'intelligenza artificiale postula il rispetto della dignità e dell'autonomia delle persone alle quali va sempre garantito un potere di supervisione sulle macchine.

Per altro verso, l'operato dell'Unione si segnala per un una peculiare attenzione per il tema della protezione dei dati personali, dimostrata anzitutto dal recepimento

---

[92] Corte e.d.u., 8 aprile 2004, *Belchev c. Bulgaria*, § 82; Corte e.d.u., 1 luglio 2003, *Suominen c. Finlandia*, § 37; Corte e.d.u., 24 luglio 2003, *Smirnova c. Russia*, § 63.

[93] Sul quale, volendo, F. ROSSI DAL POZZO – M. GIALUZ, *Commento all'art. 6*, in *Carta dei diritti fondamentali dell'Unione europea*, a cura di R. Mastroianni, O. Pollicino, S. Allegrezza, F. Pappalardo, O. Razzolini, Milano, 2017, pp. 99 ss.

[94] Cfr. THE EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Draft Ethics Guidelines for Trustworthy AI*, in https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai.

[95] Così, THE EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Draft Ethics Guidelines*, cit., p. 13.

nel *Bill of Rights* dell'Unione europea di uno specifico diritto alla protezione dei dati (art. 8 C.d.f.u.e.), ma dalla stessa giurisprudenza della Corte di Giustizia[96]. In questo quadro, non sorprende che, a livello di diritto derivato, alcune garanzie fondamentali rispetto all'utilizzo di *tools* di *risk assessment* si possono riscontrare nel recente «*data protection reform package*»[97], costituito dal regolamento 2016/679/UE (GDPR) e dalla direttiva 2016/680/UE, che sostituiscono, rispettivamente, la direttiva 95/46/CE, considerata la pietra angolare in materia di protezione dei dati personali, e la decisione quadro 2008/977/GAI. La direttiva 2016/680/UE costituisce una *lex specialis* rispetto al regolamento[98], in quanto mira a stabilire norme minime relative alla «protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti ai fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia e la prevenzione di minacce alla sicurezza pubblica»[99].

Al fine di verificare gli spazi applicativi dei *risk assessment tools* nel procedimento penale, lo strumento rilevante è dunque la direttiva 2016/680/UE.

La norma fondamentale per la materia che ci interessa è quella – che riprende una garanzia tradizionale, riconosciuta sin dall'art. 15 della Direttiva 95/46/CE – contenente il divieto di decisioni basate unicamente su trattamenti automatizzati. L'art. 11 della direttiva stabilisce infatti che «gli Stati membri dispongono che una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato sia vietata salvo che sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento e che preveda garanzie adeguate per i diritti e le libertà dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento».

Si tratta di una disposizione che, così come quella dell'art. 22 GDPR[100], ha una formulazione ambigua[101]: tutto ruota intorno all'interpretazione dell'espressione "decisione basata unicamente su un trattamento automatizzato".

Secondo una prima impostazione, tale norma vieterebbe le decisioni nelle quali «non vi è alcun coinvolgimento umano nel processo decisionale»[102]. Siffatte decisioni

---

[96] O. POLLICINO – M. BASSINI, *Commento all'art. 8*, in *Carta dei diritti fondamentali dell'Unione europea*, cit., pp. 135 ss., 157, i quali rilevano un certo attivismo della Corte di Giustizia che, sul terreno della tutela del diritto alla protezione dei dati, ha manifestato l'ambizione a ergersi, nei fatti, a Corte costituzionale dell'Unione europea.

[97] Cfr. P. DE HERT – V. PAPAKONSTANTINOU, *The New Police and Criminal Justice Data Protection Directive. A first Analysis*, in *New Journal of European Criminal Law*, 2016, n. 1, p. 7.

[98] Cfr. A. RIPOLL SERVENT, *Protecting or Processing?*, in *Privacy, Data Protection and Cybersecurity in Europe*, a cura di W.J. Shünemann e M.O. Baumann, 2017, p. 125.

[99] Cfr. art. 1, par. 1, direttiva 2016/680/UE.

[100] Con riguardo alla previsione analoga contenuta nell'art. 22 GDPR, si legga G.N. LA DIEGA, *Against the Dehumanisation of decision-Making. Algorithmic decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, in *JIPITEC*, 31 maggio 2018, pp. 18-19.

[101] Cfr. J. SAJFERT – T. QUINTEL, *Data Protection Directive (EU) 2016/680 for police and criminal justice authorities*, in COLE-BOEHM, *GDPR Commentary*, in corso di stampa, p. 10.

sono vietate, laddove producano effetti giuridici negativi oppure incidano significativamente sull'interessato: la direttiva richiede quindi di regola un intervento dell'uomo, con la specificazione che, «per aversi un coinvolgimento umano, il titolare del trattamento deve garantire che qualsiasi controllo alla decisione sia significativo e non costituisca un semplice gesto simbolico»[103]. Insomma, «*in order to escape the prohibition from Article 22GDPR or Article 11 of the Directive on Data Protection in Criminal Matters, the human has to use the machine only as decision support, whereas the final decision is taken by the human*»[104]. D'altra parte, la stessa interpretazione pare essere alla base del documento della *House of Commons* proprio sul tema "*Algorithms in decision-making*", con particolare riferimento alla portata della norma analoga dell'art. 22 GDPR: proprio facendo leva su tale interpretazione è stato ritenuto legittimo l'utilizzo del *software* HART in Inghilterra[105].

Questo è il contenuto, per così dire, minimo della disposizione. Per la verità, sembra preferibile una lettura un po' più esigente, secondo la quale, al fine di garantire un intervento effettivo dell'uomo, la stessa decisione non potrebbe basarsi esclusivamente sull'*output* di un meccanismo automatizzato[106]. Insomma, accanto all'obbligo di un intervento umano andrebbe ritenuta sussistente quella che, nel lessico processualpenalistico, chiameremmo regola di valutazione, in forza della quale l'*output* prodotto dall'IA va considerato come un mero indizio, che va sempre corroborato con altri elementi di prova[107]. Questa lettura sembra peraltro confermata dall'eccezione alla regola, contemplata dalla stessa disposizione: si ammette, infatti, che la regola possa essere derogata, a condizione che vi sia una previsione di tutele sufficienti per i diritti personali e che vi sia, *quanto meno*, un intervento umano. Di regola, pertanto, non ci si può accontentare di questo ma occorre che l'elemento cognitivo generato dall'intelligenza artificiale sia confermato da altre fonti.

Ora, quando viene in gioco la libertà personale dell'imputato, questa norma va letta assieme all'art. 5 C.e.d.u. e all'art. 6 C.d.f.u.e. e finisce per arricchire il macro-

---

[102] In tal senso, si veda il documento elaborato dal Gruppo di lavoro articolo 29 per la protezione dei dati, intitolato *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, adottate il 3 ottobre 2017 e riviste il 6 febbraio 2018, p. 23.

[103] Ancora, *Linee guida sul processo decisionale automatizzato*, cit., p. 23. Vi deve essere un «*meaningful and genuine human intervention, for instance in the form of actual oversight by a person with "authority and competence to change the decision"*» (G.N. LA DIEGA, *Against the Dehumanisation of decision-Making*, cit., p. 19).

[104] Cfr. M. BRKAN, *Do algoritms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond,* in *Electronic Journal*, gennaio 2017, p. 10. Analogamente, cfr. A. CAIA, *Commento all'art. 22 GDPR*, in *GDPR e normativa privacy. Commentario*, a cura di G.M. Riccio, G. Scorza, E. Belisario, 1ª ed., Milano, 2018, p. 223.

[105] Cfr. HOUSE OF COMMON. SCIENCE AND TECHNOLOGY COMMITTEE, *Algorithms in decision-making. Fourth Report of Session 2017–19*, in *www.parliament.uk*, 23 maggio 2018.

[106] In tal senso, in A. CAIA, *Commento all'art. 22*, cit., p. 227; G. MALGIERI – G. COMANDÉ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, vol. 7, p. 14. Così, anche G. MALGIERI – G. COMANDÉ, *Why a Right to Legibility*, cit., p. 14.

[107] In effetti, secondo le stesse *Linee guida sul processo decisionale automatizzato*, cit., p. 23, «se un essere umano riesamina il risultato del processo automatizzato *e tiene conto di altri fattori* nel prendere la decisione finale, tale decisione non sarà 'basata unicamente' sul trattamento automatizzato [corsivo aggiunto]».

diritto di accesso al giudice, di un ulteriore micro-diritto, che assume valenza centrale nella società contemporanea e che deve rappresentare un argine per gli sviluppi futuri dell'IA: l'interessato ha diritto a che sul suo *status* si pronunci un giudice in carne ed ossa, che dovrà tener conto *anche* di elementi di prova ulteriori rispetto all'*output* del *risk assesment tool*. Non vi è spazio dunque in Europa per uno scenario analogo a quello della California e del Kentucky. Sulla scorta di tale norma, l'ultima parola spetterà sempre al giudice. Anche nei casi eccezionali, nei quali il diritto dell'Unione o dello Stato membro preveda la possibilità di delegare al *software* l'adozione della decisione, dovrà esservi pur sempre la possibilità di contestare la decisione davanti a un uomo[108].

Peraltro, la norma non si limita a richiedere un intervento dell'intelligenza umana, ma contiene indicazioni significative con riferimento alla tipologia di dati che possono – o meglio non possono – essere utilizzati per la profilazione[109]. Il comma 2 esclude infatti che le decisioni automatizzate contemplate nel par. 1 possano basarsi sulle categorie particolari di dati personali, di cui all'art. 10, ossia quei dati che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche o l'appartenenza sindacale, nonché i dati genetici, quelli biometrici o relativi alla salute o, ancora, i dati relativi alla vita sessuale della persona o all'orientamento sessuale, a meno che non vi siano «misure adeguate a salvaguardia dei diritti, delle libertà e dei legittimi interessi dell'interessato». È ben vero che non si tratta del divieto assoluto che aveva suggerito il Parlamento in prima lettura[110], ma è comunque un presidio significativo.

Vi è, infine, un divieto assoluto posto dal par. 3, con riguardo alla profilazione basata sui dati appena ricordati che porti alla discriminazione di persone fisiche[111].

---

[108] V. J. SAJFERT – T. QUINTEL, *Data Protection Directive*, cit., p. 10.

[109] La definizione di profilazione è fornita dall'art. 3, n. 4: «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o *prevedere* aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, *il comportamento*, l'ubicazione o gli spostamenti di detta persona fisica». Si tratta di una forma automatizzata di trattamento, effettuato su dati personali e finalizzato a *valutare aspetti personali* relativi a una persona fisica: viene impiegata per effettuare previsioni su persone usando dati provenienti da varie fonti per dedurre qualcosa su quella persona in base alle qualità di altre persone che appaiono statisticamente simili. Tre sono i momenti rilevanti: la raccolta dei dati; l'analisi automatizzata per individuare correlazioni; l'applicazione della correlazione a una persona fisica per effettuare previsioni su comportamenti futuri (*Linee guida sul processo decisionale automatizzato*, cit., p. 7-8).

[110] Nell'art. 9, comma 2-*ter*, si prevedeva: «È vietata in tutti i casi la profilazione che, intenzionalmente o meno, dia luogo a discriminazioni basate su razza, origine etnica, opinioni politiche, religione o convinzioni personali, appartenenza sindacale, genere o orientamento sessuale, o che comporti, intenzionalmente o meno, misure aventi tali effetti discriminatori» (P7_TA(2014)0219, *Trattamento dei dati personali ai fini di prevenzione di reati. Risoluzione legislativa del Parlamento europeo del 12 marzo 2014 sulla proposta di direttiva del Parlamento europeo e del Consiglio concernente la tutela 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, e la libera circolazione di tali dati (COM(2012)0010 – C7-0024/2012 – 2012/0010(COD)*).

[111] Cfr. J. SAJFERT – T. QUINTEL, *Data Protection Directive*, cit., p. 11.

Da questa ricostruzione emerge come vi sia, già oggi, tanto a livello di Consiglio d'Europa, quanto a livello di Unione europea, una serie di regole che consentono, per un verso, di salvaguardare il ruolo dell'intelligenza umana nei processi decisionali e, per l'altro, di vietare alla radice l'impiego di *tools* che si basino sul trattamento di dati sensibili e che siano suscettibili di condurre a discriminazioni.

7. Una prospettiva futuribile anche in Italia?

Qualche considerazione conclusiva merita di essere fatta sull'Italia. Evidentemente, nel corso del procedimento penale, il giudice è chiamato a svolgere diversi giudizi predittivi sulla pericolosità dell'imputato. Tra i più delicati, quelli in materia cautelare, relativi alla sussistenza dei *pericula libertatis*, nonché, all'esito del giudizio, quello in fase di determinazione della pena, sulla capacità a delinquere del reo[112].

Come noto, il legislatore si è molto impegnato nel definire i criteri di tali giudizi prognostici.

Per un verso, le norme dell'art. 274 c.p.p. – in particolare alla lett. *c* – e dell'art. 133, comma 2, c.p., fissano dei temi di prova per le parti (art. 187 c.p.p.); per altro verso, delineano dei tracciati motivazionali per il giudice. Al di là delle differenze lessicali, gli elementi da valutare sono parzialmente coincidenti: occorre verificare le modalità e circostanze del fatto, nonché il carattere o la personalità del prevenuto, desunta «da comportamenti o atti concreti o dai suoi precedenti penali» (secondo l'art. 274, comma 1, lett. *c*, c.p.p.), oppure dai precedenti penali e giudiziari e, in genere, dalla condotta e dalla vita del reo, antecedenti al reato; dalla condotta contemporanea o susseguente al reato; infine, dalle condizioni di vita individuale, familiare e sociale del reo (art. 133, comma 2, c.p.). Si tratta di dati oggettivi attinenti alla vita del soggetto.

È altrettanto noto che il legislatore processuale italiano ha sempre voluto tener fuori la scienza dai giudizi predittivi. Le ragioni sono sostanzialmente due.

L'una ha a che fare con la tutela della presunzione di innocenza[113], ma non convince: la difficoltà è ammettere giudizi predittivi di pericolosità di un soggetto presunto innocente durante il processo (e conosciamo bene l'infinito dibattito sulla legittimità dell'esigenza cautelare di prevenzione della pericolosità[114]); ma una volta ammessi, non ha senso limitare gli strumenti cognitivi del giudice.

---

[112] Che questa vada valutata come prognosi dei futuri comportamenti dell'agente discende da una lettura costituzionalmente orientata della disposizione dell'art. 133, comma 2, c.p. (v. per tutti G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di Diritto Penale. Parte generale*, 7ª ed., Milano, 2017, p. 706).

[113] In tal senso, si legga la posizione risalente di G. VASSALLI, *Il potere discrezionale del giudice nella commisurazione della pena*, in *Conferenze. Primo Corso di perfezionamento per uditori giudiziari*, Torino, 1958, ora in *Scritti giuridici*, vol. I, t. 2, Milano, 1997, p. 1334.

[114] Si leggano, per tutti, G. AMATO, *Individuo e autorità nella disciplina della libertà personale*, Milano, 1967, p. 380; F. BRICOLA, *Politica criminale e politica penale dell'ordine pubblico (a proposito della l. 22 maggio 1975, n. 152)*, in *La questione criminale*, 1975, p. 248 s.; V. GREVI, *Libertà personale dell'imputato e Costituzione*, Milano, 1976, p. 44 s.; G. ILLUMINATI, *La presunzione d'innocenza dell'imputato*, Bologna, 1979, p. 42 ss.

La verità è che la ragione più profonda del divieto in parola risiede nel fatto che il legislatore italiano ha sempre avuto assai poca fiducia nella scienza psicologica, criminologica, che mira a esplorare il foro interno dell'interessato. In quest'ottica, il tradizionale divieto di perizia criminologica – scolpito nell'art. 314, comma 2, c.p.p. 1930 e ribadito nell'art. 220, comma 2, del codice Vassalli – presenta una stretta connessione con l'art. 188 c.p.p.; non a caso, per giustificare tale limite, Franco Cordero scriveva: «sono troppi i soi-disants macchinisti dell'anima ed è meglio che non mettano piede nel processo»[115]. In definitiva, è questa la ragione per la quale il legislatore del 1988 ha confermato un divieto giudicato anacronistico dalla dottrina, fin dagli anni Sessanta[116], e dalla stessa Corte costituzionale[117].

La conseguenza di questa «strumentazione processuale *inadeguata*» è duplice: «da un lato, la tendenza a sfumare l'accertamento del carattere e della personalità, in genere, del reo nel regno delle intuizioni e delle impressioni d'atmosfera, che oltre a essere difficilmente traducibili per iscritto, non riescono agilmente verificabili; dall'altro lato, una certa inerzia dei giudici, sia nel motivare, sia nell'affrontare l'indagine della seconda parte dell'art. 133 c.p.»[118].

Ora, mi sembra che i *risk assessment tools* non siano finalizzati specificamente a scandagliare il foro interiore dell'interessato per accertare «il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche»: non è quindi scontato che il loro utilizzo vada qualificato come una vera e propria perizia criminologica e che rientri quindi nell'ambito di applicazione del divieto dell'art. 220, comma 2, c.p.p. Secondo una certa impostazione scientifica – peraltro, come si è visto, non indiscussa nel panorama americano[119] – sono in grado di leggere in modo apparentemente più efficace di quanto riesca a fare l'intelligenza umana degli indici fattuali esteriori per effettuare dei giudizi predittivi. Il punto allora non è tanto vietarli *a priori*, facendosi scudo di una norma basata su una presunzione assoluta di inattendibilità di altri metodi scientifici; l'alternativa, conviene tenerlo bene a mente, tanto in materia di libertà personale, quanto in materia di quantificazione della pena, è una valutazione affidata «all'intuito del giudice, quando non addirittura al suo incontrollabile *arbitrium*»[120].

---

[115] Così, con la consueta efficacia, F. CORDERO, *Codice di procedura penale commentato*, Torino, 1990, p. 264.

[116] Basti far riferimento a F. BRICOLA, *La discrezionalità nel diritto penale*, Milano, 1965, p. 116 e G. VASSALLI, *Criminologia e giustizia penale*, in *Scritti giuridici in onore di Alfredo De Marsico*, a cura di G. Leone, vol. II, Milano, 1960, p. 581. Più recentemente, si legga G. VARRASO, *La prova tecnica*, in *Trattato di procedura penale*, diretto da G. Spangher, vol. II, *Prove e misure cautelari*, t. 1, *Le prove*, a cura di A. Scalfati, Torino, 2009, p. 242-243.

[117] Il riferimento è a Corte cost., 24 giugno 1970, n. 124, la quale non esclude che «la diffidenza verso la perizia psicologica sia discutibile di fronte allo sviluppo degli studi moderni sulla psiche ed è auspicabile che la norma sia aggiornata».

[118] Così, F. BRICOLA, *La discrezionalità*, cit., p. 116 (la prima frase si trova nella nota 231).

[119] V. C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 17, secondo i quali «*the predictive accuracy of risk assessment algorithms remains an open question*».

[120] In tal senso, con riferimento alla valutazione di cui all'art. 133 c.p., G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, 7ª ed., Bologna, 2014, p. 802.

Si tratta allora di verificare, anzitutto, la validità del modello matematico che sta al fondo dello strumento, ma soprattutto la genuinità dei dati che vengono utilizzati dal *tool*: la benzina che alimenta qualsiasi sistema basato sull'IA è costituita dai dati ed è fondamentale non solo la quantità, ma anche la qualità di questi. Ove il meccanismo lavori su dati imprecisi o inconferenti il rischio di produrre un *output* inattendibile (o peggio, discriminatorio) è elevatissimo. Siccome gli sviluppatori dei sistemi non dispongono dei dati è fondamentale che l'apporto interdisciplinare già nella fase della progettazione dell'algoritmo.

In secondo luogo, occorre garantire la trasparenza del processo valutativo effettuato dallo stesso strumento e la conseguente possibilità di contestare l'affidabilità dell'*output*. A tal fine risulterebbe forse preferibile coinvolgere direttamente le agenzie pubbliche e la comunità scientifica per costruire dei sistemi trasparenti e rispettosi dei canoni fissati dalla Carta etica europea per l'uso dell'intelligenza artificiale[121].

Da ultimo, occorre capire, alla luce della Costituzione, quali sono i limiti e gli spazi della giustizia penale nei quali i *risk assessment tools* potranno trovare applicazione. Occorre evitare il rischio che attraverso questi strumenti si apra la strada a una forma inaccettabile di determinismo penale, per cui dal diritto penale del fatto – sancito dall'art. 25, comma 2, Cost. – si passi a un inaccettabile diritto penale del profilo d'autore, nel quale la pericolosità di un soggetto viene desunta esclusivamente dagli schemi comportamentali e dalle decisioni assunte in una determinata comunità nel passato. Ovviamente, questo sarebbe contrario al principio di individualizzazione del trattamento sanzionatorio, desumibile dall'art. 27, comma 1 e 3, Cost.[122], nonché, del canone di individualizzazione del trattamento cautelare, ricavabile dagli artt. 13 e 27, comma 2, Cost.[123]. Occorre allora chiedersi se da tale vincolo di individualizzazione si possa desumere una vera e propria regola di esclusione della valutazione di pericolosità fondata su meccanismi che si basano su generalizzazioni di condotte di soggetti diversi dall'interessato. Il tema è delicato ed è difficile fornire una risposta in termini generali, indipendentemente dalla struttura del dispositivo; in fondo, le stesse massime di esperienza che vengono impiegate dall'intelligenza umana per effettuare giudizi predittivi di responsabilità nascono da generalizzazioni di esperienze di altri soggetti. Quel che cambia, nel caso dell'IA è la quantità di fattispecie passate e di informazioni che vengono prese in considerazione e, naturalmente, i criteri di valutazione ed elaborazione dei dati. Come si è detto, appare dunque essenziale garantire la trasparenza dell'algoritmo e coinvolgere i giuristi nella sua creazione, affinché i dati impiegati siano affidabili e pertinenti alla valutazione di pericolosità e i criteri di valutazione non siano irragionevoli o discriminatori.

---

[121] Cfr. anche i suggerimenti prospettati, soprattutto per evitare le discriminazioni basate sulla razza e sulla condizione sociale, C. DOYLE – C. BAINS – B. HOPKINS, *Bail Reform*, cit., p. 17.

[122] Secondo Corte cost., 2 aprile 1980, n. 50, «l'"individualizzazione" della pena, in modo da tenere conto dell'effettiva entità e delle specifiche esigenze dei singoli casi, si pone come naturale attuazione e sviluppo di principi costituzionali».

[123] V., in particolare, Corte cost., 21 luglio 2010, n. 265, § 5 (considerato in diritto).

Detto questo, non credo che un'attività umana tra le più importanti per la comunità – perché ha ad oggetto i beni personali più preziosi, tanto per il singolo, quanto per la società – come il giudizio predittivo di pericolosità compiuto dal giudice penale, possa escludere *a priori* – in forza di una presunzione assoluta e aprioristica di inattendibilità – l'ausilio (non, sia chiaro, la sostituzione) della scienza e della tecnologia. Proprio quella statistica *bayesiana* che sta alla base del *decision making* in tanti ambiti della nostra vita: pensiamo alle scelte effettuate, con risvolti talvolta tragici, nell'ambito della medicina o dei trasporti aerei. Deve essere ben chiaro che, se tali strumenti funzionano e rispettano i diritti fondamentali, sarà difficile tenerli fuori dal perimetro della giustizia penale: già si stanno sviluppando nella fase della prevenzione e in quel segmento più deformalizzato del procedimento, rappresentato dalle indagini preliminari; non v'è dubbio che si svilupperebbero anche nella fase dell'esecuzione. Sarebbe assurdo avere un giudice che giudica sempre e solo sulla base del suo intuito. Peraltro, non va sottovalutato che in prospettiva potrebbe essere lo stesso sviluppo tecnologico a rendere accessibile direttamente al giudice il *risk assessment*: nel momento in cui – lo si dice provocatoriamente – il *tool* di valutazione del rischio potrà essere scaricato con una *app* sul telefonino, si sarà ampliato il sapere comune del giudice (ossia il patrimonio culturale dell'uomo medio) che segna lo spartiacque tra scienza privata e competenze specifiche.

Ad ogni modo, la garanzia fondamentale è rappresentata dalla centralità del giudice, nel segno tracciato dall'art. 11 della direttiva 2016/680/UE e dell'art. 8, d.lgs. 18 maggio 2018, n. 51, che ha dato attuazione in Italia alla direttiva: gli strumenti di intelligenza artificiale possono aiutare il giudice, ma mai sostituirlo. In fondo, la Corte costituzionale, nella bellissima sentenza n. 124 del 1970, ha scorto alla base del divieto di perizia criminologica (anche) la preoccupazione «che lo studio della personalità dell'imputato possa venir compiuto *solo da chi abbia presente anche il carattere afflittivo e intimidatorio della pena* [corsivo aggiunto]»[124]. Come si è detto, non appare ragionevole escludere strumenti diversi (potenzialmente utili a rendere più giusto e meno arbitraria la valutazione), ma è fondamentale – come ha riconosciuto la stessa Corte suprema del Wisconsin nel caso *Loomis*[125] – che l'ultima parola spetti sempre e comunque al giudice, il quale deve mantenere la *sua* autonomia ed evitare quello che viene chiamato l'«*automation complacency*»[126] o *automation bias*, che si verifica nel processo decisionale perché gli esseri umani hanno la tendenza a ignorare o a non cercare informazioni che contraddicono la soluzione generata dal computer che è accettata come corretta[127].

In conclusione, vi sono elementi che inducono a ritenere che l'intelligenza artificiale possa offrire un ausilio all'uomo anche nei giudizi predittivi da compiere

---

[124] Corte cost., 24 giugno 1970, n. 124, cit.

[125] Cfr. *supra*, § 2.

[126] R. PARASURAMAN – D. H. MANZEY, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, in *Human Factors*, 2010, p. 381

[127] M.L. CUMMINGS, *Automation Bias in Intelligent Time Critical Decision Support Systems*, *Paper presented to the American Institute for Aeronautics and Astronautics First Intelligent Systems Technical Conference*, 2004, p. 1, consultabile a questo *link*.

nell'ambito della giustizia penale. Non dobbiamo rifiutarla *a priori* e neanche accettarla supinamente. Dobbiamo imparare a farne un uso critico, trovando un equilibrio tra la positiva riduzione dell'arbitrio e i pericoli di un'inaccettabile visione deterministica che cancellerebbe le garanzie di libertà[128]. Certamente è un compito arduo, ma la strada è tracciata dalla cornice di garanzie definite a livello costituzionale ed europeo.

[128] Sulle opportunità e i rischi dell'IA, cfr., da ultimo, L. FLORIDI – J. COWLS – M. BELTRAMETTI – R. CHATILA – P. CHAZERAND – V. DIGNUM – C. LUETGE – R. MADELIN – U. PAGALLO – F. ROSSI – B. SCHAFER – P. VALCKE – E. VAYENA, *An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, accessibile a questo *link*, p. 2 ss.