INTELLIGENZA ARTIFICIALE E GIUSTIZIA: NELLA CORNICE DELLA CARTA ETICA EUROPEA, GLI SPUNTI PER UN'URGENTE DISCUSSIONE TRA SCIENZE PENALI E INFORMATICHE.

di Serena Quattrocolo
(*Professore ordinario di diritto processuale penale, Università del Piemonte orientale*)

SOMMARIO: 1. L'iniziativa della CEPEJ – 2. Gli obiettivi della Carta – 3. I principi della Carta; 3.1. Il rispetto dei diritti fondamentali; 3.2. La non discriminazione; 3.3Qualità e sicurezza; 3.4 Trasparenza, imparzialità e fairness; 3.5 Il controllo dell'utente – 4. Gli scenari peculiari della giustizia penale

1. Il 4 dicembre 2018 è stata emanata, nell'ambito del Consiglio d'Europa, la 'Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti'. Il documento è stato stilato dalla CEPEJ, Commissione europea per l'efficacia della giustizia, istituita nel 2002 per iniziativa del Comitato dei Ministri del Consiglio d'Europa, con lo scopo di monitorare e misurare la qualità dei sistemi giudiziari dei Paesi membri. La commissione, dotata di numerose divisioni e gruppi di lavoro, predispone ogni due anni una relazione sullo stato dei sistemi di giustizia nazionale, stimandone la qualità, i tempi, l'indipendenza, la composizione di genere, sulla base di parametri come il budget, il personale, l'utenza, l'organizzazione, il numero delle decisioni. La più recente relazione, basata sui dati 2016, è stata pubblicata lo scorso ottobre e fotografa in modo articolato lo 'stato di salute' della giustizia nei diversi ordinamenti nazionali[1].

L'attenzione 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, in tutti i settori delle istituzioni, anche quelle giudiziarie, è nota. Già nel marzo del 2018 veniva pubblicato l'interessante studio dal titolo *Algorithms and Human Rights*[2], nella cui sezione dedicata a *fair trail* e *due process* si anticipavano alcune delle preoccupazioni cui la nuova Carta intende rispondere.

Infatti, con diverse velocità e con specifiche peculiarità, tutti i settori della giustizia, civile, amministrativa e anche penale, nell'ultimo decennio sono entrati in contatto – o, piuttosto, si sono scontrati – con l'inesorabile filtrare, dalla realtà d'oltreoceano verso quella europea, di sistemi computazionali che garantiscono la trattazione automatizzata, per le più varie finalità, di enormi quantità di dati, con

[1] Reperibile all'indirizzo https://www.coe.int/en/web/cepej/special-file-publication-2018-edition-of-the-cepej-report-european-judicial-systems-efficiency-and-quality-of-justice-
[2] DGI (2017)12, reperibile sul sito ufficiale del Consiglio d'Europa.

tempi e costi ridottissimi[3]. Per quanto riguarda specificamente la giustizia penale, l'impatto è avvenuto, al momento, per lo più sul piano probatorio, ove le potenzialità dell'analisi dei dati generati dagli strumenti digitali divenuti di uso comune si sono innanzitutto mostrate attraverso nuove forme di captazione occulta di conversazioni e di molti altri dati riservati, con una capacità intrusiva nella sfera di riservatezza delle persone senza precedenti[4] e, per questo, difficilmente inscrivibile all'interno della cornice normativa già esistente, nella maggior parte dei Paesi europei. Molti ordinamenti hanno recentemente affrontato, più o meno efficacemente, la questione dei captatori informatici[5] per cercare di ristabilire – pur nell'incertezza che la recente rivoluzione digitale ha determinato rispetto a concetti classici, quali il 'domicilio' e la 'riservatezza' – la protezione di diritti inviolabili degli individui, sanciti sia da molte Costituzioni nazionali, sia dalla Convenzione europea dei diritti dell'uomo. Inoltre, sempre sul piano probatorio, l'impiego di dati conoscitivi generati in maniera del tutto automatica può determinare altre gravi lesioni di garanzie fondamentali del processo equo[6] e, in particolare, della parità delle armi[7], che la stessa Carta sembra cogliere e voler scongiurare. Prima di scendere però nei contenuti del documento, pare opportuno premettere quali sono gli obiettivi in esso riposti dalla CEPEJ.

2. La Carta etica europea costituisce, nella sua forma attuale, un emblematico esempio di *soft law*. Proviene da un autorevole organismo internazionale ed è rivolto in maniera indistinta a soggetti pubblici e privati, a vario titolo coinvolti nella realizzazione e nell'utilizzo di strumenti di intelligenza artificiale che analizzano dati e decisioni giudiziarie, come recita il breve preambolo del documento. La Carta si rivolge anche «ai legislatori chiamati a stabilire una cornice normativa all'interno della quale tali strumenti vanno sviluppati, verificati e utilizzati». L'affermazione,

[3] U. Pagallo – M. Durante, *The Philosophy of Law in an Information Society*, in L. Floridi (a cura di), *The Routledge Handbook of Philosophy of Information*, New York 2016, 396 ss.

[4] *Ex multis*, M. Daniele, *La prova digitale nel processo penale*, in *Riv. Dir. Proc.* 2011, p. 288; M. Pittirutti, Digital evidence e procedimento penale, Torino 2017, *passim*.

[5] Si veda il comprensivo studio commissionato dal LIBE Committee del Parlamento europeo, *Legal Frameworks for hacking by Law Enforcement: Identification, Evaluation and Comparison of Practices* (reperibile sul sito ufficiale del Parlamento Europeo). Con riferimento alla recente riforma italiano v. P Bronzo, *Intercettazione ambientale tramite captatore informaticco: limiti di ammissibilità, uso in altri processi e divieti probatori*, in G. Giostra - R. Orlandi (a cura di), *Nuove norme in tema di intercettazioni*, Torino 2018, 235 ss.; M. Torre, *Il captatore informatico: nuove tecnologie investigative e rispetto delle regole processuali*, Milano 2017, *passim*

[6] E. Van Buskirk – V.T. Liu, *Digital Evidence: Challenging the Presumption of Reliability*, in *Journal of Digital Forensic Practice* 2006, 19 ss.

[7] U. Pagallo – S. Quattrocolo, *The impact of AI in criminal Law, and its Twofold Procedures*, in W. Barfield- U. Pagallo (a cura di), *Research Handbook on the Law of Artificial Intelligence*, Cheltenam, 2018, 388 ss.

evidentemente priva di termini imperativi, riconosce come indiscusse la necessità e l'urgenza che si proceda in tal senso, offrendo la Carta come paradigma essenziale per tale attività. Infatti, per la Commissione, l'impiego di metodi computazionali per il rafforzamento della efficacia della giustizia deve essere incoraggiato. Questo mi pare il primo punto essenziale da mettere in luce nell'individuazione degli obiettivi del documento. All'interno dell'insuperabile cornice costituita dalla Convenzione europea dei diritti dell'uomo, della Convenzione sulla protezione dei dati personali e delle specifiche garanzie enunciate nel testo della Carta, la CEPEJ mostra apertura – e non ostracismo – verso strumenti che possono determinare un aumento dell'efficienza complessiva dei sistemi di giustizia. Dalla lettura dell'appendice esplicativa (Annex I) emerge, però, che il valore aggiunto del ricorso a strumenti di IA non necessariamente è – o deve essere – quello proposto e reclamizzato dai produttori del *software* stesso: ad esempio, attraverso la rapida costruzione di correlazioni tra quintilioni di dati tratti da precedenti decisioni giudiziarie, un sistema di intelligenza artificiale *non è* in grado di spiegare il ragionamento giuridico, ma può soltanto, attraverso i modelli elaborati sulla base di quelle correlazioni, esprimere la verosimiglianza che il giudice propenda per una decisione analoga a quella già adottata in circostanze simili[8] (senza escludere il rischio di correlazioni errate). Questa importante precisazione si aggancia opportunamente al terzo principio stabilito dalla Carta che, come si vedrà tra breve, cristallizza la necessità di un costante approccio multidisciplinare nello sviluppo, nella verificazione e nell'applicazione di strumenti computazionali, affinché essi possano davvero rappresentare un'evoluzione nell'efficienza della giustizia e non un elemento di rischio per le garanzie fondamentali degli individui.

In questi termini, ben emerge dall'intervista rilasciata dal segretario esecutivo della Commissione in occasione della pubblicazione della Carta[9], il senso ultimo dell'iniziativa, ovvero l'affermazione di una consapevolezza, non ancora sufficiente, in tutti gli attori coinvolti sulla scena della giustizia, riguardo alle sfide lanciate dalla rivoluzione digitale[10] e del ruolo essenziale che i diritti fondamentali già elaborati all'interno del Consiglio d'Europa devono giocare, come cornice imprescindibile nell'incontro tra i due mondi, delle scienze dure e delle scienze sociali, apparentemente inconciliabili.

[8] V. Annex 1, p. 27.
[9] Stéphan Leyemberger, reperibile alla homepage del sito istituzionale www.coe.int/cepej
[10] Interessante la lettura proposta da A. Garapon – J. Lassègue, *Justice digitale*, Parigi 2018,1 9 ss., che vede nel digitale una rivoluzione grafica che – sulla scia di quelle che in precedenza hanno segnato la storia, come ad esempio il comparire dell'alfabeto greco – sta producendo un impatto epocale sulla comunicazione e sui suoi riflessi.

3. Riassunti così gli obiettivi della Carta – che, tra i molti documenti già adottati in vari contesti[11], rappresenta un primo modello formulato in articoli – è opportuno riflettere brevemente su ciascuno dei cinque principi ivi cristallizzati.

3.1. Il primo principio riguarda il **rispetto dei diritti fondamentali**, *assicurando che la progettazione e l'applicazione dei sistemi di IA e dei relativi servizi siano compatibili con i diritti fondamentali*. La formula può suonare declamatoria, ma va sottolineato con forza che questo documento, pur non vincolante, si rivolge non solo e non principalmente agli Stati (come spesso accede per gli atti del Consiglio d'Europa), ma per lo più a singoli individui e, in particolare, a operatori del settore privato, come sono solitamente gli sviluppatori di *software*, ingegneri, matematici, magari analisti, comprensibilmente estranei all'articolato sistema delle garanzie fondamentali costruito all'interno del Consiglio d'Europa. La nota esplicativa sottolinea come i due principali testi di riferimento siano la Convenzione europea dei diritti dell'uomo e la Convenzione n. 108 sulla protezione dei dati personali. In questi due strumenti si individuano alcune garanzie che debbono ispirare l'attività di creazione del *software*, secondo un orientamento che la nota felicemente definisce *human-rights-by-design*, con una terminologia ben nota a chi si occupa di *privacy*. In concreto, ciò significa concepire e 'addestrare' l'IA – sia essa uno strumento di risoluzione del contenzioso, un elemento di supporto al *decision-making* giurisdizionale, o un mezzo di divulgazione al pubblico delle decisioni giudiziarie – in modo da evitare violazioni dei diritti fondamentali connessi all'amministrazione della giustizia e, in particolare: il diritto di accesso alla giurisdizione; il diritto al processo equo, nelle sue articolazioni essenziali del contraddittorio e della parità delle armi; il principio di legalità; l'indipendenza della magistratura e, in particolare, dei giudici nell'esercizio del potere decisorio. Sullo sfondo di queste opportune raccomandazioni si intravedono, infatti, scenari che vanno attentamente analizzati, poiché anche le forme più semplici e diffuse di strumenti computazionali applicati alla giustizia possono ingenerare effetti 'di sistema', dall'impatto eclatante. L'appendice di accompagnamento opportunamente riflette sulle possibili conseguenze delle *open data policies* recentemente adottate da un buon numero di Paesi del Consiglio d'Europa. Premesso che l'accessibilità digitale a tutte le sentenze pronunciate in un ordinamento (*open data*) – con possibilità di analisi e ricerca attraverso parole chiave e raggruppamenti di parole – non equivale e non sostituisce il tradizionale principio di pubblicità delle decisioni giudiziarie, diversamente garantito, è opportuno domandarsi quale effetto produrrà tale diffusa e illimitata accessibilità sul valore del 'precedente',

[11] Noto e fondamentale è l'impegno costante dell'organizzazione IEEE, Insitute of Elettrical and Electronic Engeneers, che da anni si occupa di fornire una cornice etica delle applicazioni dell'IA.

soprattutto negli ordinamenti che non sono su di esso basati. La forte correlazione tra un certo gruppo di fattori e una determinata decisione giudiziaria, rivelata da un sistema computazionale (non necessariamente molto sofisticato), può determinare uno 'scivolamento normativo' verso una maggiore vincolatività del precedente, imponendo un onere motivazionale rafforzato al giudice che se ne voglia distaccare? La domanda è di eccezionale attualità[12], soprattutto perché un sistema *open data* non attentamente calibrato rischia di confondere il piano della quantità delle decisioni con quello della qualità delle medesime, rovesciando il rapporto tra giurisdizioni inferiori e superiori, alle quali spetta, per lo più, il ruolo di nomofilachia. In questo senso, il tema della indipendenza del giudice riemerge significativamente, al di là della tradizionale funzione del principio nella tutela della magistratura giudicante rispetto agli altri poteri dello Stato, in particolare quello esecutivo[13].

Tale riflessione incidentale rende evidente come l'articolo 1 faccia riferimento a garanzie da tempo incorporate nella maggior parte delle costituzioni nazionali e nella Convenzione europea, e arricchite quindi da un'articolata giurisprudenza della Corte di Strasburgo (sulla quale in sede di commento 'a prima lettura' non v'è modo di soffermarsi), i cui significati e le cui implicazioni sono note solo al giurista 'specializzato' e inaccessibili a chi abbia competenze diverse. In tale rilievo si sintetizza la sfida maggiore rappresentata dalla Carta, che intende promuovere l'instaurazione di una collaborazione biunivoca tra aree del sapere al momento non connesse tra loro, prive di un linguaggio comune, forse perché, fino ad ora, non ve n'era stata una precisa ragione… Il primo articolo della Carta fissa invece un nuovo obiettivo, che giustifica e sollecita la creazione di strumenti di reciproca comprensione e interazione.

3.2. Il secondo principio, di **non discriminazione**, *specificamente fa divieto di creare o accentuare discriminazioni tra gruppi e individui*. Al di là del significato 'tradizionale' che il canone ha in numerosi ambiti normativi, il problema della possibile discriminazione assume un significato particolare in relazione all'elaborazione automatizzata dei dati. Se è vero, infatti, come sottolinea l'Appendice esplicativa, che i sistemi computazionali sono particolarmente adatti a individuare l'esistenza di eventuali discriminazioni, questi stessi – basati su modelli o algoritmi che partendo da una serie di input elaborano degli output specifici – sono soggetti al rischio di c.d. *implicit bias*. Plurimi sono i livelli ai quali il rischio si può verificare. In primo luogo, se l'*input* non è completamente neutro[14], l'*output* dell'interrogazione rischia di essere

[12] Si veda la forte critica di A. Garapon – J. Lassègue, cit., 279 s.

[13] Sempre attuale il quadro tracciato da M. Chiavario, *sub art. 6 Cedu*, S. Bartole - B. Conforti - G. Raimondi (a cura di), *Commentario alla Convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padova 2002, 182.

[14] Ad esempio, l'indirizzo di abitazione non è un dato neutro, perché può suggerire – pure

influenzato da un pregiudizio, che può portare alla discriminazione di singoli individui o di gruppi sociali. In secondo luogo, l'algoritmo – che è concepito e interpretato da un umano – può banalmente riprodurre ingiustificati preconcetti sociali… Anche su questo aspetto la letteratura è ormai enorme e sarebbe difficile offrirne una sintesi esaustiva[15].

Pure sotto questo profilo, però, appare evidente come la necessità di evitare l'accentuazione di discriminazioni esistenti o, addirittura, la creazione di nuove, implichi una strettissima sinergia tra esperti di intelligenza artificiale e modelli computazionali e gli studiosi dei processi e delle interazioni sociali, per leggere oltre i dati analizzati e individuare eventuali effetti patologici scaturenti dal ricorso a *data set* solo apparentemente neutri.

3.3 Il terzo principio, di **qualità e sicurezza,** 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.* Come già ampiamente anticipato, 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.

La prima parte della previsione, però, si concentra specificamente sulla sicurezza dei dati giudiziari elaborati attraverso sistemi computazionali. Oltre, infatti, ai rischi già segnalati nel paragrafo precedente, la scelta dei dati da immettere nel processo di elaborazione implica l'attenta verifica dell'affidabilità della fonte e della integrità del dato, per evitarne modificazioni, accidentali o strumentali. A tale scopo, l'intero processo deve essere tracciato e verificabile *ex post*: il contenuto e il significato della decisione processata non deve poter essere alterato in alcun passaggio. Per la medesima ragione, si legge nella 'glossa' del principio n. 3, occorre che i modelli e gli algoritmi su cui si fonda elaborazione siano operati (e custodi) in ambienti sicuri, evitando rischi per la loro integrità e intangibilità.

Il rispetto del principio n. 3 si collega direttamente con il profilo della protezione del segreto industriale e commerciale del *software*: il disvelamento del suo funzionamento può comportarne la divulgazione e, quindi, la riproduzione da parte di concorrenti commerciali. Come opportunamente sottolinea anche l'appendice di accompagnamento, la questione non si è posta al momento, in Europa, con la stessa intensità con cui si è presentata negli Stati Uniti, ove per una diversa regolamentazione

erroneamente – l'appartenenza a un determinato gruppo etnico o sociale, che potrebbe risultare discriminato dall'applicazione dai risultati dell'interrogazione del sistema.

[15] Per un'interessante e aggiornata panoramica v. T.P. Woods, *The Implicit Bias of the Implicit Bias Theory, in Drexel Law Review* 2017, 631 ss.

normativa e per un uso certamente più massiccio dei modelli computazionali nei sistemi di giustizia, si sono verificati alcuni clamorosi casi di diniego di accesso della difesa ai codici che regolano un sistema digitale predittivo della pericolosità sociale, usato dalle Corti, sulla base della tutela del *trade secret*.[16] La necessità di un bilanciamento di valori tra la salvaguardia degli interessi commerciali e il diritto di difesa rappresenta il trait d'union tra il terzo principio e quello successivo, in una fitta rete di interrelazioni che contraddistingue tutto il documento.

3.4. Il quarto principio enunciato dalla Carta, infatti, appare tanto familiare, nel suo frequente ricorrere, quanto arcano nella sua effettiva applicazione. Affermando il principio 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. In questa formulazione si riflette la preoccupazione già sopra espressa in merito al complesso rapporto tra protezione della proprietà intellettuale e del *trade secret* e la necessità di osservare, capire, criticare i processi computazionali utilizzati. Se la preoccupazione vale per tutti i settori della pubblica amministrazione, essa assume un valore ancor più spiccato rispetto alla giustizia e, nell'ambito di questa, particolarmente con riguardo al processo penale, in cui sono in gioco i più elevati beni giuridici, come la libertà personale. Come già anticipato, l'interesse della giustizia deve prevalere nel bilanciamento con gli interessi privati degli sviluppatori del *software* e ciò può avvenire soltanto se tutto il ciclo del modello, dal design alla sua applicazione quotidiana sono ispirati ai tre fattori sopra enunciati, della trasparenza, imparzialità e *fairness*. Tuttavia, non si può negare che l'istanza di segretezza dei codici-sorgente sia del tutto priva di rilievo, non solo con riguardo agli

[16] Il caso - noto come Compas (Correctional Offender Managment Profiling for Alternative Sanctions), nome del *software* predittivo, o Loomis, nome dell'imputato - è stato deciso nel 2016 dalla Corte suprema del Wisconsin (su cui v. *Harvard Law Review* 2017, pp. 1530-1537). Il ricorrente, condannato alla pena di sei anni di reclusione e cinque anni di 'sorveglianza' lamentava la violazione del *fair trail* per via dell'impiego da parte della Corte delle risultanze del Compas *risk assessment*, il cui funzionamento non era stato disvelato per tutelare il *trade secret*. La Corte Suprema statale, nel rigettare il ricorso, ha tuttavia sottolineato la cautela che i giudici devono impiegare nell'utilizzo di tali strumenti 'predittivi'. Sebbene il *software* avesse goduto di valutazioni potenzialmente positive in letteratura (pur nell'affermato bisogno di ulteriore validazione: v. T. Brennan – W. Dietrich – B. Ehret, *Evaluating the Predictive Validity of the Compas Risks and Needs Assessment System,* in *Criminal Jusice and Behaviour* 2009, 21 ss.), uno studio di Angwin et alii pubblicato dalla ONG americana ProPublica ha mostrato la scarsa rilevanza criminogena di alcuni fattori utilizzati dal modello. È bene tuttavia segnalare che le conclusioni dello studio diffuso da ProPublica sono state fortemente criticate (v. A.W. Flores – K. Bechtel – C.T Lowenkamp, *False Positives, False Negatives and False Analysis: A Rejoinder to «Machine Bias: There is Software used across the Country to Predict Future Criminals. And it is biased against the Blacks»*, in *Federal Probation* 2016).

interessi commerciali, ma anche per esigenze interne al procedimento penale, quando si tratti di strumenti utilizzati a fini investigativi.[17]

A prescindere dalle soluzioni tecniche che caso per caso possono essere utilizzate per rispondere alle specifiche esigenze che si presentino nel singolo procedimento, la soluzione generale e principale auspicata dalla Carta è la completa trasparenza tecnica[18], accompagnata da un'esplicazione del processo computazionale in linguaggio accessibile e chiaro. È noto, infatti, che nonostante l'enfasi riposta nel concetto di 'trasparenza'[19], questa non rappresenti di per sé la soluzione di tutti gli squilibri generati dall'impiego di dati prodotti automaticamente: 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'[20]. Come anticipato, ciò assume nel settore della giustizia un significato particolare, legato alla salvaguardia del principio fondamentale della pubblicità del processo decisionale e, in particolare, con riguardo alla valutazione della prova: a fronte della necessità – espressamente prevista da alcuni ordinamenti, come quello italiano – che la motivazione della sentenza dia conto della valutazione di attendibilità operata dal giudice rispetto a ciascuna prova, la 'trasparenza algoritimica' non è di per sé sufficiente a fornire al giudice, ai destinatari della decisone e all'opinione pubblica l'effettiva comprensione del processo che ha portato a generare la prova digitale, lasciando quindi avvolto nell'incertezza anche il giudizio sulla sua attendibilità.

A questo scopo, utile opzione collaterale è individuata dalla Carta nella creazione di autorità indipendenti che possano verificare e certificare *a priori*, e periodicamente, i modelli impiegati nei servizi della giustizia. Si tratta di un suggerimento che completa la triade dei principi enunciati nell'art. 4, proponendo uno strumento ispirato proprio a imparzialità e *fairness*. Il documento non si spinge oltre sul punto, lasciando ai suoi destinatari 'istituzionali' di cogliere e modellare il suggerimento in base alle caratteristiche di ciascun ordinamento. È importante, tuttavia, aver attirato l'attenzione sul tema, che ad oggi, nell'Europa continentale, sembra appartenere ancora all'area del futuribile, anziché a quella dell'esistente: la previsione e l'instaurazione delle garanzie suggerite, per essere efficace, deve essere

[17] In questo senso, volendo, S. Quattrocolo, *Equità del processo penale e* automated evidence *alla luce della convenzione europea dei diritti dell'uomo*, in *Rev. italo-española derecho procesal* 2018 (2), p. 12.
[18] M. Hildebrandt, *Profile transparency by design? Re-enabling double contingency*, in M. Hildebrandt – K. De Vries (a cura di), *Privacy, Due Process and the Computational Turn*, Londra 2013, 221-246
[19] "Transparency seems to have replaced legitimacy as the core value of data protection": P. De Hert - S. Gutwirth, *Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power*, in E. Cleas - A. Duff - S. Gutwirth (a cura di), *Privacy and the Criminal Law*, Anversa 2006, 80.
[20] A. Koene - H.Webb - M Patel, *First UnBias Stakeholders workshop*, 11 in https://unbias.wp.horizon.ac.uk/dissemination/publications/

preventiva rispetto all'ingresso sul mercato di *software* che forniscono servizi per la giustizia, soprattutto nella forma di strumenti di assistenza alla decisione.

3.5 Il quinto principio enunciato dalla Carta è il **controllo da parte dell'utente**, che *esclude un approccio prescrittivo* dell'impiego dell'intelligenza artificiale, *garantendo che gli utilizzatori agiscano come soggetti informati, nel pieno controllo delle loro scelte*. Infatti, gli strumenti computazionali e i sistemi di intelligenza artificiale devono accrescere l'autonomia decisionale dell'utente, non ridurla, come afferma espressamente la 'glossa'.

Benché nella spiegazione del principio si colga un concetto non univoco di 'utente' – talvolta riferito al professionista della giustizia che utilizza il servizio, talaltra al soggetto destinatario della decisione 'automatizzata' o basata sulle risultanze del servizio stesso – il canone racchiude in sé numerose implicazioni essenziali.

Partendo proprio dall'ultima considerazione espressa nel testo, la condizione minima per il rispetto di questo principio è l'esistenza e la diffusione, tra gli operatori della giustizia, di adeguata letteratura esplicativa e di un dibattito scientifico che coinvolga anche i giuristi. Dato questo substrato minimo, la tutala dell'autonomia dell'utente richiede però numerosi altri capisaldi, modulati in relazione alle specifiche finalità che il servizio automatizzato intende fornire.

In primo luogo, con riguardo all'elaborazione, *open data*, delle decisioni giudiziarie, il testo riprende la raccomandazione, già formulata, circa la necessità di chiarire espressamente il margine di vincolatività del precedente, individuando gli spazi di autonomia correlati alle peculiarità di ciascun singolo caso, posta la necessaria possibilità che l'operatore risalga – e possa analizzare – (al)le decisioni dalle quali è derivato il risultato automatizzato. In secondo luogo, quando il sistema computazionale sia invece rivolto all'ausilio dell'attività decisionale del giudice, entra in gioco una gamma diversa di tutele. Data anche qui la necessità di specificare la natura vincolante o meno della 'decisione digitale', il destinatario della medesima deve avere il diritto all'assistenza legale e di acceso al giudice, nonché di essere informato preventivamente dell'impiego di un ausilio automatizzato alla decisione giudiziale, in modo da poter contestare tale possibilità davanti ad un giudice, nel quadro delle garanzie stabilite dall'art. 6 Cedu.

Come già in precedenza osservato, si tratta di profili che toccano l'essenza dei sistemi di giustizia nazionali, che una diffusione superficiale e disattenta dei molti strumenti computazionali già esistenti nel mondo potrebbe minare alla base. E pur rinviando ad altra sede la riflessione più approfondita sulle singole tematiche qui soltanto accennate, occorre sottolineare come l'appendice di accompagnamento alla Carta sviluppi attentamente, in particolare tre profili critici, uno dei quali

espressamente collegato con la giustizia penale. Innanzitutto, si pone in dubbio tanto che i sistemi di intelligenza artificiale possano riprodurre (e anticipare) il ragionamento giuridico (basandosi su mere correlazioni tra gruppi di parole e decisioni assunte in precedenza e esprimendo tassi elevati di probabilità, senza tuttavia riprodurre in alcun modo la complessità del ragionamento giuridico, almeno così come tradizionalmente inteso in Europa) – quanto che, analogamente, essi siano in grado spiegare retrospettivamente il comportamento del giudice: infatti, nella massa di correlazioni evidenziate dall'intelligenza artificiale, solo alcuni fattori sono stati effettivamente causativi della decisione ed essi dipendono dall'esistenza di una cornice interpretativa che può essere compresa solo attraverso il filtro di diverse scienze sociali[21]. Certo, tali informazioni addizionali potrebbero esser tradotte in *inputs* per la trattazione automatica, ma occorrerebbe un passaggio ulteriore, che non può basarsi sui dati conoscitivi offerti da un sistema *open data* (spesso, l'unico dato a disposizione è il nome del giudice estensore e/o del presidente…).

4. Più specificamente, poi, con riguardo alla giustizia penale, lo studio annesso alla Carta si sofferma intanto sull'inadeguatezza del termine 'giustizia predittiva' ormai ampiamente in uso anche in Europa. Sulla base di numerose e durature esperienze nord-americane, l'uso di strumenti 'predittivi' per la prevenzione dei reati e per la valutazione del rischio di recidivanza o di pericolosità sociale[22] emerge sempre più frequentemente nel dibattito sulle prospettive della giustizia penale. Al di là del criticabile scivolamento che questi strumenti 'predittivi' possono determinare tra le sfere della prevenzione e della repressione del reato[23], si segnala il rischio che pure gli ordinamenti europei, per lo più orientati da tempo alla individualizzazione della sanzione penale subiscano, anche in ragione dell'impiego di tali modelli computazionali, un arretramento verso dottrine deterministiche.

Mentre il dibattito è stato inizialmente incentrato sulla attendibilità delle teorie (bayesiane) predittive integrate in tali *softwares*, recentemente sono emerse, nel contesto americano, la scarsa rappresentatività dei dati spesso utilizzati per il *risk assessment* e l'insufficiente falsificazione sul piano scientifico del metodo proposto da ciascun modello. Sotto il primo profilo, studi citati dalla stessa appendice alla Carta hanno messo in luce come molti dei dati impiegati per la valutazione di rischio di

[21] Annex I, cit., 39.

[22] Per una interessante e aggiornata panoramica v. N. Scurich, *The case against categorical risk estimates*, in *Behavioral Science Law* 2018, 1 ss.

[23] J.A.E. Vervaele, *Surveillance and Criminal Investigation: Blurring of Thresholds and Boundaries in the Criminal Justice System*?, in S. Gutwirth – R. Leenes – P. De Hert (a cura di), *Reloading Data Protection*, Heidelberg 2014, 114 ss.

comportamenti violenti in realtà non abbiano alcun rilievo sul piano criminogeno, ma sottolineino soltanto dei fattori di maggiore vulnerabilità sociale, che hanno dato luogo, con l'uso diffuso di tali *softwares* nelle corti americane, ad un'accentuazione delle discriminazioni già esistenti (v. supra § 3.2). Sotto il secondo profilo invece, una recente decisione della corte suprema del District of Columbia, sezione minorile[24], avendo ribadito che l'impiego di specifici *softwares* predittivi, anche ai fini della valutazione del *risk assessment* è subordinato al noto *Daubert test*, ha sottolineato l'imprescindibile condizione che ogni metodo sia sottoposto a riproduzione e falsificazione, attraverso la pubblicazione scientifica e la *peer review*, escludendo nel singolo caso di specie l'utilizzabilità dei risultati del *risk assessment* digitale.

Sottolinea l'appendice che, essendo il processo decisionale umano regolato da una gerarchia di priorità, l'inserimento nel modello computazionale di dati riferiti al gruppo sociale di appartenenza può significare che il comportamento passato di un certo gruppo possa incidere sul destino dell'imputato, senza che il *software* possa cogliere, appunto, quell'innata gerarchia di priorità che regola l'agire umano. E quale valore la valutazione digitale debba assumere nella decisione del giudice in ordine alla colpevolezza e alla quantificazione della pena dipende da ogni singolo ordinamento. Quanto e come il giudice possa distaccarsi dalla valutazione predittiva diviene questione di particolare importanza, soprattutto negli ordinamenti in cui la magistratura è una carica elettiva. Non si nasconde, tuttavia, la preoccupazione che ciò possa assumere un rilevo sul piano disciplinare e della responsabilità civile anche negli ordinamenti in cui la magistratura – almeno quella giudicante – gode di un'indipendenza maggiore.

È evidente come tali argomenti rappresentino, in questa sede, solo degli 'appunti' per la riflessione futura, che dovrà toccare davvero gli aspetti più essenziali dei sistemi di giustizia europei, opportunamente inseriti nella cornice delle garanzie stabilite dal Consiglio d'Europa, oggi arricchita proprio dalla Carta che si è voluta qui presentare. Dalla pregnanza di questi argomenti discendono comunque le importanti conclusioni raccolte nell'Appendice n. 2 alla Carta: stabiliti quattro livelli di approccio consigliato ai diversi metodi computazionali da applicare ai servizi della giustizia (da incoraggiare; attuabili, ma con significative precauzioni metodologiche; futuribili, previo approfondimento scientifico; uso da prendersi in considerazione con le più estreme riserve possibili), l'impiego di strumenti computazionali predittivi nel processo penale e l'approccio normativo basato sulla quantità dati sono stati collocati proprio nell'ultima categoria.

[24] Superior Court of the District of Columbia – Family Court (Juvenile and Neglected Branch), T.K., 15.3.2018, giudice R.D. Okun, inedita.

Ciò sollecita non un radicale rigetto, ma un autentico dibattito scientifico e la massima attenzione agli strumenti che potrebbero penetrare nei nostri ordinamenti non a seguito di una necessaria, attenta riflessione, ma sulla base di un mero incontro tra offerta e domanda di mercato.