



12 LUGLIO 2023

Il nuovo volto della discrezionalità
giudiziaria: prospettive e pericoli a
partire dalla giurisprudenza americana
sui risk assessment tools impiegati nel
sentencing

di Andrea Valsecchi

Dottorando di ricerca in Diritto costituzionale
Università degli Studi di Milano



Il nuovo volto della discrezionalità giudiziaria: prospettive e pericoli a partire dalla giurisprudenza americana sui risk assessment tools impiegati nel sentencing*

di Andrea Valsecchi

Dottorando di ricerca in Diritto costituzionale
Università degli Studi di Milano

Abstract [It]: Negli Stati Uniti, l'impiego di strumenti algoritmici di previsione del rischio di recidiva (cd. *risk assessment tools*) ha dato origine a un'ampia casistica giurisprudenziale, precedente e successiva al noto caso *Loomis*, chiamata a sindacare l'abuso di discrezionalità nel momento in cui il giudice assume la propria decisione (anche) sulla base del risultato algoritmico. Muovendo da siffatta casistica, ancora pressoché inesplorata, il presente contributo intende indagare il rinnovato perimetro della discrezionalità giudiziaria e il mutamento del suo esercizio dinanzi all'algoritmo predittivo.

Title: The new face of judicial discretion: perspectives and dangers from U.S. case law on risk assessment tools in sentencing

Abstract [En]: In the United States, the use of algorithmic tools for predicting the risk of recidivism (so-called risk assessment tools) has given rise to an extensive case law, preceding and following the well-known *Loomis* case, called upon to review the abuse of discretion when the judge makes his decision (also) on the basis of the algorithmic result. From such case law that is still almost unexplored, the essay aims to investigate the renewed perimeter of judicial discretion and the change in its exercise before the predictive algorithm.

Parole chiave: *Risk assessment tools*, algoritmi predittivi, giustizia penale, discrezionalità del giudice, abuso di discrezionalità

Keywords: Risk assessment tools, predictive algorithms, criminal justice, judicial discretion, abuse of discretion

Sommario: 1. Il perimetro dell'indagine: l'esercizio della discrezionalità di fronte all'algoritmo predittivo. 2. I *risk assessment tools* impiegati nel *sentencing*. 2.1. L'evoluzione storica della valutazione del rischio di recidiva negli Stati Uniti. 2.2. Breve descrizione del funzionamento dei *tools*. 2.3. L'acquisizione del risultato algoritmico da parte del giudice nella fase di determinazione in concreto della pena. 3. L'ampia casistica giurisprudenziale sui *risk assessment tools*: non solo *Loomis*. 4. (*Segue*) Critiche sistematiche e ordinamentali all'utilizzo nel *sentencing* degli strumenti predittivi. 5. (*Segue*) I primi argomenti a favore della legittimazione dell'uso dei *tools* da parte del giudice. 6. (*Segue*) Il peso dell'algoritmo nel percorso discrezionale del giudice: il labile confine tra *reliance* e *consideration*. 7. Esternalizzazione della discrezionalità giudiziaria e opacità dell'algoritmo: una insidiosa "relazione biunivoca". 8. La motivazione come argine *all'abuse of discretion*? 9. Uno sguardo sull'Europa: i primi embrionali tentativi di regolare l'intelligenza artificiale all'interno dei sistemi giudiziari. 10. Alcuni spunti conclusivi: la formazione dei giudici e la ricerca di un nuovo principio processuale: *algorithmia novit curia*.

* Articolo sottoposto a referaggio.

1. Il perimetro dell'indagine: l'esercizio della discrezionalità giudiziaria di fronte all'algorithmo predittivo

Quando si discute di giustizia algoritmica, intesa come la ricerca di un efficientamento della macchina giudiziaria mediante l'impiego di algoritmi e tecnologie automatizzate per supportare o financo sostituire le decisioni del giudice, si citano spesso le parole pronunciate nel 2017 dal *Chief Justice* della Corte suprema degli Stati Uniti, John G. Roberts Jr., il quale, alla domanda se riuscisse a prevedere un giorno in cui macchine intelligenti guidate dall'intelligenza artificiale (IA) sarebbero state in grado di assistere i tribunali nell'accertamento dei fatti o, persino, nel «*judicial decision making*», la sua risposta apparve particolarmente netta: «*[i]t's a day that's here, and it's putting a significant strain on how the judiciary goes about doing things*»¹.

Nonostante siano parole, in effetti, particolarmente efficaci, esse tendono la mano a ricerche e osservazioni che limitano il proprio campo di indagine alle possibili violazioni delle garanzie del giusto processo (in particolare, a titolo di esempio, il diritto di difesa, il diritto di accesso al giudice, il diritto a un ricorso effettivo)².

I lodevoli sforzi della dottrina lasciano tuttavia in ombra ciò su cui invero le parole del Giudice Roberts parrebbero voler mettere in guardia: la giustizia algoritmica sta mettendo a dura prova (*significant strain*) il modo stesso di esercitare la funzione giurisdizionale, le modalità attraverso cui il giudice opera, gli strumenti giuridici che accompagnano la decisione giudiziaria.

La «*tremendous revolution*» – come definita dal Giudice – non conosce quindi solo un mutamento di paradigma dal punto di vista di coloro che vengono raggiunti da decisioni in tutto o in parte assunte da sistemi algoritmici ma, al contrario, determina un significativo cambio di prospettiva dal lato di chi tali sistemi è chiamato a utilizzare, nella specie il giudice.

Per discutere giuridicamente di un tema così complesso si porrà l'attenzione sulla giurisdizione statale statunitense che da anni sperimenta l'impiego di sistemi algoritmici in sede processuale, in particolare i cd. *risk assessment tools* (o *risk and needs assessment tools*, RNA), *software* utilizzati per valutare il rischio di recidiva di un soggetto sottoposto a giudizio penale. La dottrina, sia italiana sia straniera, si è

¹ L'intervista integrale è disponibile su *YouTube* al presente [link](#).

² Sul tema specifico dell'impatto dell'IA sulle garanzie del giusto processo la letteratura, in Italia, è molto varia. Si segnalano, tra gli altri, C. CASTELLI - D. PIANA, *Giusto processo e intelligenza artificiale*, Maggioli Editore, Santarcangelo di Romagna, 2019; S. QUATTROCOLO, *Equità del processo penale e automated evidence alla luce della convenzione europea dei diritti dell'uomo*, in *Revista Ítalo-Española de Derecho Procesal*, pp. 107-123; S. QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, in *BioLaw Journal – Rivista di BioDiritto*, n. 1/2019, pp. 135-144; G. CONTISSA - G. LASAGNI - G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e LA: alla ricerca di un rimedio effettivo*, in *Diritto di internet*, n. 4/2019, pp. 619-634. Con riferimento alla letteratura straniera si v. *ex multis*, J. ULENAERS, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, in *Asian J Law Econ*, vol. 11(2), 2020; K. TERZIDOU, *The Use of Artificial Intelligence in the Judiciary and Its Compliance with the Right to a Fair Trial*, in *Journal of judicial administration*, vol. 31 (3), 2022, pp. 154-168; A. Z. HUQ, *Constitutional Rights in the Machine Learning State*, in *Cornell Law Review*, Vol. 105, 2020.

principalmente concentrata sul noto caso *Loomis*³ (si cfr. *infra*, § 5) e su pochi altri casi emblematici, e ha ben monitorato – come anticipato – il rispetto delle garanzie costituzionali afferenti al giusto processo; ancora poco indagata è, invece, la cospicua casistica giurisprudenziale che, almeno dal 2008 sino ad oggi, è stata chiamata a sindacare la legittimità dell'utilizzo dei *risk assessment tools* nella fase di *sentencing* del processo penale, la fase processuale in cui il giudice, all'esito della dichiarazione di colpevolezza del reo, è chiamato a determinare in concreto la pena da applicare allo stesso⁴. L'oggetto di siffatto sindacato, al contrario di quanto si creda, solo in casi sporadici è rappresentato dal rispetto della *due process clause*⁵; piuttosto, nella maggior parte delle decisioni l'attenzione dei giudici, sollecitati dalle doglianze delle parti, si concentra sull'abuso di discrezionalità del giudice (*abuse of discretion*) nel momento in cui questi fonda la propria decisione (anche) sulle risultanze di un algoritmo predittivo.

L'esercizio della discrezionalità è quindi quasi sempre al centro degli arresti giurisprudenziali che si occupano di *tools* algoritmici: pertanto, il presente contributo intende misurarsi con la giustizia algoritmica con un approccio diverso, da un lato assumendo, per le ragioni indicate, il giudice come epicentro del fenomeno in atto e, dall'altro, tentando di vedere come l'esercizio del potere muti dinanzi all'algoritmo predittivo. La prospettiva del lavoro non potrà dunque che muovere dalle numerose sentenze dei giudici americani da cui è possibile trarre alcune indicazioni sul rinnovato perimetro della discrezionalità del giudice, ancora poco ricercato e che invece sarà oggetto di un forte cambiamento man mano che strumenti dei IA verranno implementati nella fase decisionale del procedimento giudiziario.

Dopo una sommaria indicazione dell'evoluzione degli strumenti di predizione del rischio di recidiva impiegati nell'ordinamento giudiziario americano, l'indagine muoverà da alcuni cenni essenziali sul

³ Con riguardo alla letteratura straniera si cfr., in generale sui rischi e le opportunità dei *risk assessment tools*, *ex multis*, C. SLOBOGIN, *Principles of Risk Assessment: Sentencing and Policing*, in *Ohio State Journal of Criminal Law*, vol. 15, 2018, pp. 583-596; J. VILASENOR - V. FOGGO, *Artificial intelligence, due process, and criminal sentencing*, in *Mich. St. L. Rev.*, 2020, pp. 295-354; R. BERK *et al.*, *Fairness in Criminal Justice Risk Assessments: The State of the Art*, in *Sociological Methods & Research*, 50 (1), pp. 3-44; Z. XU, *Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities*, in *Applied Artificial Intelligence*, vol. 36 (1), 2022. Nello specifico sul caso *Loomis* e le sue ricadute sul giusto processo si cfr. H. LIU - C. LIN - Y. CHEN, *Beyond State v Loomis: artificial intelligence, government algorithmization and accountability*, in *International Journal of Law and Information Technology*, 2019, vol. 27, pp. 122-141; *Wisconsin Supreme Court Requires Warnings before Use of Algorithmic Risk Assessment in Sentencing*, in *Harvard Law Review*, 130, 5, 2017, pp. 1530 ss.; S. GOEL - R. SHROFF - J. L. SKEEM - C. SLOBOGIN, *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment*, dicembre 2018, disponibile in [SSRN](#). In Italia, il tema dei *risk assessment tools*, trainato dal caso *Loomis*, ha catturato l'attenzione soprattutto dei penalisti e dei processualpenalisti; si cfr. *ex multis*, v. V. MANES, *L'oracolo algoritmico e la giustizia penale*, in *DisCrimen*, 15 maggio 2020; G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale*, 11 novembre 2020; F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. e uomo*, 29 settembre 2019; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. Dir. Proc. Pen.*, vol. 4, 2019; S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, 4, 2019, 1748 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti e Europa*, in *Dir. Pen. Cont.*, 29 maggio 2019; D. ZINGALES, *Risk assessment: una nuova sfida per la giustizia penale? La pericolosità criminale al vaglio algoritmico delle probabilità nell'esperienza statunitense*, in *Dir. Pen. e uomo*, 9 dicembre 2021.

⁴ B. L. GARRET - J. MONAHAN, *Judging Risk*, in *California Law Review*, vol. 108, 2020, p. 441.

⁵ Come noto riconducibile al V e al XIV emendamento alla Costituzione degli Stati Uniti.

funzionamento di siffatti *tools*, la cui comprensione è un prerequisito necessario per affinare lo sguardo sulle problematiche giuridiche che essi pongono.

Queste ultime verranno discusse, come detto, esaminando le pronunce dei giudici degli Stati federati che fanno correntemente uso dei *risk assessment tools*, al fine di metterne in luce alcuni aspetti decisivi per comprendere il modo in cui l'esercizio della discrezionalità giudiziaria sta assumendo caratteri affatto nuovi e diversi, meritevoli di venire valorizzati da una prospettiva costituzionale e ordinamentale.

La giurisprudenza americana suggerirà infine dei percorsi di riflessione sull'impiego dell'algoritmo, sul ruolo del giudice e sull'esercizio del suo potere, in particolar modo sul margine discrezionalità a lui assegnato, tenendo a mente in ogni caso che spesso, come si è autorevolmente affermato in tempi risalenti, «potere e discrezionalità vengono [...] a identificarsi»⁶.

2. I risk assessment tools impiegati nel sentencing

La giustizia penale statunitense è stata – ed è tutt'oggi – la più importante testimone dell'utilizzo diffuso di strumenti algoritmici di predizione del rischio di recidiva. Se prima tali sistemi venivano impiegati esclusivamente al di fuori del processo strettamente inteso, oggi negli Stati Uniti quasi ogni fase procedimentale ospita l'utilizzo di sistemi predittivi: le attività di polizia di ricerca del reato, la scelta di misure cautelari e delle misure alternative alla detenzione, la fase di esecuzione della pena e, infine, il cd. *sentencing*⁷. Per fornire un primo assaggio di come il *sentencing* sia influenzato dall'impiego di sistemi predittivi, in uno studio apparso nel 2017 all'esito di un progetto condotto dal *National Centre for State Courts (NCSC)* si indicava come nell'ultimo decennio almeno diciotto Stati negli USA avessero consentito l'uso dei *tools* nella fase del *sentencing*⁸.

L'idea di utilizzare regole standardizzate per (parzialmente) razionalizzare la discrezionalità giudiziaria in sede decisionale non è nuova: già negli anni '70 il giudice Marvin Frankel pose al pubblico il problema dell'eccessiva discrezionalità dei giudici americani, in un saggio che influenzò per anni il dibattito sul tema⁹.

Secondo Frankel, il potere arbitrario del giudice nel *sentencing* doveva essere fortemente limitato e le decisioni dei giudici, per poter essere considerate legittime, avrebbero dovuto basarsi su «criteri oggettivi

⁶ C. MORTATI, *Note sul potere discrezionale*, 1936, ora in *Problemi di diritto pubblico nell'attuale esperienza costituzionale repubblicana*, Raccolta di scritti, vol. III, 1972, p. 1009.

⁷ D. KEHL - P. GUO - S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, in *Responsive Communities Initiative, Berkman Klein Center for Internet & Society*, Harvard Law School, 2017, pp. 13 ss.

⁸ NATIONAL CENTRE FOR STATE COURTS, *Use of risk and needs assessment information in state sentencing proceedings*, settembre 2017, p. 3. In generale, negli ultimi studi disponibili si è stimato che nel mondo ci siano più di quattrocento *tools* utilizzati nell'ambito della giustizia penale, si cfr. B. L. GARRET - J. MONAHAN, *Judging Risk*, cit., p. 450.

⁹ M. FRANKEL, *Criminal Sentences: Law Without Order*, Hill and Wang, New York 1973.

e oggettivamente verificabili»¹⁰. La sua proposta era quella di creare una tabella o un elenco di fattori che potessero contenere «ove possibile, una forma di valutazione numerica o comunque oggettiva»¹¹ e una commissione che analizzasse le sentenze e formulasse delle linee guida standardizzate per ricondurre la discrezionalità giudiziaria su binari più prevedibili. Il libro di Frankel fu uno dei testi più influenti del periodo e diede avvio a una serie di studi su larga scala relativi ai fattori che possono influire sulle decisioni dei giudici (fattori che un recente studio definisce con il nome di “rumore”¹², e che comprendono, tra gli altri, il momento della giornata in cui il giudice assume la propria decisione; la vittoria o la sconfitta della squadra di *football* locale preferita; il giorno del compleanno del singolo condannato o dello stesso giudice)¹³.

Sebbene il seguito delle stimolazioni di Frankel sia stato incerto e travagliato ed esuli dalle finalità del presente contributo¹⁴, ciò che pare utile sottolineare è il rinnovato interesse – spinto dalla imperturbata evoluzione tecnologica – verso strumenti che, come le linee guida vincolanti, sono in grado di *orientare* la decisione del giudice, nella specie quello penale. In questo senso, tra le applicazioni che più stanno riscuotendo fortuna spiccano per diffusione e rilevanza i già menzionati *risk assessment tools*.

2.1. L'evoluzione storica della valutazione del rischio di recidiva negli Stati Uniti

L'evoluzione degli strumenti di predizione del rischio di recidiva ha conosciuto diverse fasi e la letteratura è concorde nell'individuare almeno quattro distinte generazioni¹⁵, ancorché tale classificazione, a fronte della continua e incessante trasformazione dei *tools*, non sembra poter rinunciare oggi, come vedremo, a un'ulteriore fase¹⁶, proprio per la distanza dei sistemi di più recente sviluppo rispetto ai precedenti.

La prima generazione di strumenti di predizione del rischio di recidiva, apparsa negli Stati Uniti nella prima metà del secolo scorso¹⁷, non contemplava sistemi artificiali ma si fondava sull'analisi di esperti e periti clinici che esprimevano un proprio giudizio professionale sul rischio riscontrato.

¹⁰ *Ivi*, p. 11.

¹¹ *Ivi*, p. 114.

¹² D. KAHNEMAN - O. SIBONY - C. R. SUNSTEIN, *Noise: A Flaw in Human Judgement*, trad. it. *Rumore: un difetto del ragionamento umano*, Milano, 2021.

¹³ *Ivi*, p. 23.

¹⁴ Le sollecitazioni di Frankel portarono all'approvazione, nel 1984, del *Sentencing Reform Act*, una riforma che condusse all'istituzione della *US Sentencing Commission* che avrebbe approvato le linee guida vincolanti per i giudici nel procedimento per la determinazione della pena. Le linee guida subirono poi diverse critiche – in particolare perché impedivano al giudice di vagliare attentamente le circostanze di ciascun caso – e vennero revocate dalla Corte suprema nel 2005, la cui deliberazione eliminò il carattere vincolante delle linee guida, che assunsero quindi un valore solo consultivo. Per i dettagli si veda, ancora, D. KAHNEMAN - O. SIBONY - C. R. SUNSTEIN, *Noise*, cit., pp. 25 ss.

¹⁵ Sulla lunga storia dei *tools* si rinvia più diffusamente a M. HAMILTON, *Risk-needs assessment: constitutional and ethical challenges*, in *American Criminal Law Review*, vol. 52, 2015, pp. 236 ss. nonché a D. KEHL - P. GUO - S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, cit., pp. 9 ss.

¹⁶ Condividono tale impostazione B. L. GARRET - J. MONAHAN, *Judging Risk*, cit., p. 451.

¹⁷ D. KEHL - P. GUO - S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, cit., p. 8.

A partire dagli anni '70 vennero introdotti metodi artificiali per “misurare” il rischio, con l'utilizzo di macchine in grado di analizzare una serie di fattori sulla cui base produrre la propria analisi di rischio. La seconda generazione di *tools* comprende, quindi, strumenti meccanici che si basano su fattori *statici* (es. il sesso, l'età al momento dell'arresto, i precedenti penali: sono i cd. *historical factors*) che tuttavia risultarono presto inadatti in quanto consideravano solo fattori non “controllabili” dal soggetto sottoposto all'analisi¹⁸.

Per ovviare al problema iniziarono a venire implementati nelle macchine anche fattori *dinamici* (terza generazione) quali l'impiego lavorativo, dipendenze da sostanze, l'età vista come fattore che muta nel tempo e così via (sono i cd. *criminogenic factors*), che permettevano di focalizzare l'attenzione non solo sul versante retributivo ma anche su quello rieducativo del reo¹⁹.

I *tools* di quarta generazione, invece, utilizzano le tecniche dei precedenti sistemi (quindi l'analisi di fattori sia statici sia dinamici) ma migliorano nell'*accuracy* adottando, tramite procedure algoritmiche, un approccio più completo rispetto ai singoli fattori di rischio che, caso per caso, vengono in rilievo per misurare il rischio di recidiva di un singolo soggetto²⁰.

Sebbene alcuni autori ricomprendano nella quarta generazione anche i sistemi più moderni, sembra opportuno considerare un'ultima e distinta fase di evoluzione – attualmente in corso – nella quale i *tools* cominciano a essere implementati attraverso tecniche di intelligenza artificiale, e in particolare di *machine learning*²¹: detti sistemi sollevano inevitabilmente, quanto al loro impiego da parte dei giudici, interrogativi autonomi e specifici dei quali si dirà nel prosieguo.

2.2 Breve descrizione del funzionamento dei *tools*

In virtù delle diverse fasi generazionali che i *tools* hanno conosciuto, all'interno delle diverse giurisdizioni statali sono presenti, oggi, strumenti algoritmici di ogni tipo; descrivere il funzionamento di ognuno di essi è impresa ardua e comunque non essenziale ai nostri fini.

¹⁸ Un esempio di *tool* di seconda generazione è Static-99, il più utilizzato per il *sexual recidivism*, che contiene dieci fattori statici, cinque dei quali ricollegabili alla storia criminale del reo, si cfr. M. HAMILTON, *Risk-needs assessment: consituitional and ethical challenges*, cit., p. 237.

¹⁹ Esempi di *tools* di terza generazione sono HCR-20, lo strumento più utilizzato per la previsione della recidiva violenta, e Level of Service Inventory-Revised (LSI-R), che incontreremo nel corso del lavoro; si cfr, ancora, M. HAMILTON, *Risk-needs assessment: consituitional and ethical challenges*, cit., p. 238.

²⁰ In M. HAMILTON, *Risk-needs assessment: consituitional and ethical challenges*, cit., p. 238, viene fatto rientrare nella quarta generazione anche il noto algoritmo COMPAS, che tuttavia, secondo la classificazione qui proposta, è considerabile come uno strumento di quinta generazione.

²¹ Il *machine learning* è definibile come la capacità dei sistemi di IA di «*sophisticatedly perform the task without any intervention of human beings on the basis of learning and constantly increasing experience to understand the problem complexity and need for adaptability*», v. J. ALZUBI - A. NAYYAR - A. KUMAR, *Machine Learning from Theory to Algorithms: An Overview*, in J. Phys.: Conf. Ser., 1142, 2018.

Ciò che invece rileva è tentare di comprendere il funzionamento dei sistemi maggiormente evoluti per due ordini di ragioni: innanzitutto non vi è dubbio che con il passare del tempo saranno tali strumenti (della quarta o della quinta generazione, a seconda della classificazione che si accoglie) a venire impiegati in misura prevalente; in secondo luogo, sono gli strumenti che rispetto alle finalità della presente indagine presentano maggiori criticità e spunti di riflessione per ragionare del rinnovato perimetro della discrezionalità giudiziaria.

Uno dei *tools* più evoluti, il più conosciuto e il più studiato²², è COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), strumento prodotto dalla società Equivant (già Northpoint). Prendendo le mosse dalla guida pratica realizzata dalla società produttrice del *software* (*Practitioner's Guide to COMPAS Core*)²³, è possibile descriverne brevemente il funzionamento, la cui comprensione ci permetterà di estendere la riflessione anche agli altri *tools*, con il rischio di qualche inevitabile ma innocua approssimazione.

La Guida realizzata da Equivant non ha una valenza meramente descrittiva e la sua importanza pratica si evince anche dal fatto che almeno una volta è stata utilizzata da un giudice per respingere le doglianze di un imputato che lamentava la violazione del *due process* per mancanza di trasparenza²⁴.

Una – seppure embrionale – comprensione del funzionamento di COMPAS ha anche lo scopo di rideterminare il perimetro del discorso sull'intelligenza artificiale nel processo penale, che molto spesso ha travalicato i confini della realtà sfiorando discorsi avveniristici lontani da ciò che la dinamica processuale americana offre.

Occorre subito rilevare che non vi è alcuna indicazione sul personale che ha progettato il *software*, e dunque non è possibile ricavare il *background* di competenze alla base della progettazione di COMPAS.

Inoltre, nella Guida si ammette che il sistema è frutto di un compromesso tra il tentativo di comprendere il maggior numero di fattori criminogeni e la facilità di utilizzo dello stesso. In altre parole, per garantire l'efficacia ed efficienza dello strumento si è scelto quali fattori criminogeni considerare nell'analisi e quali, invece, sacrificare²⁵.

Il funzionamento del *software* COMPAS si divide in due “indici” o “scale”: le scale di rischio (*risk scales*)²⁶ e le scale di bisogno (“*need scales*”). Le prime hanno uno scopo predittivo (sono cioè progettate per indicare una previsione del rischio di recidiva), le seconde invece hanno il fine di misurare le tendenze criminogene del soggetto e identificare meglio il programma di intervento²⁶. Ciascuna scala COMPAS è stata costruita

²² K. B. FORREST, *When machines can be judge, jury, and executioner: justice in the age of artificial intelligence*, World Scientific Publishing Co. Pte. Ltd., Singapore, 2021, p. 65.

²³ La guida completa, realizzata dalla società produttrice, è reperibile online al seguente *link*: [Guida Compas.pdf](#).

²⁴ *State v. Loomis*, 881 N.W.2D 749 (Wis 2016) §§ 54 e 69.

²⁵ Guida COMPAS, p. 2.

²⁶ Guida COMPAS, p. 12.

sulla base di una serie di fattori comportamentali e psicologici di grande rilevanza per il calcolo della recidiva²⁷. I dati che l'algoritmo analizza sono autonomamente raccolti dai dipartimenti che ne fanno utilizzo, anche mediante la somministrazione al reo di un questionario²⁸.

Le scale di rischio che COMPAS è in grado di analizzare sono diverse, ma ai nostri fini è sufficiente considerare in generale come queste sono costruite. Gli *input* della scala includono: il *background* criminoso (arresti precedenti, condanne passate ecc.), la situazione professionale o familiare, l'eventuale consumo di droghe, l'età al momento della valutazione COMPAS e l'età del primo arresto. Tutti fattori che, secondo la teoria criminologica accolta da COMPAS, sono decisivi per la previsione della recidiva²⁹.

L'analisi algoritmica di tali fattori viene svolta con tecniche di *machine learning*³⁰ (anche se nella guida non si specifica quali) e produce un indice di recidiva da 1 a 10 (da 1 a 4 *Low*, da 5 a 7 *Medium*, da 8 a 10 *High*). A causa dell'utilizzo di fattori standardizzati di predizione (cioè frutto di una elaborazione statistica dei dati) non è da sottovalutare lo scenario in cui il sistema produca dei risultati controintuitivi: la guida stessa specifica che l'algoritmo potrebbe determinare un rischio alto di recidiva per un soggetto che non ha commesso alcun reato (perché, in ipotesi, è giovane e vive in un contesto socioeconomico classificato come pericoloso) e un rischio basso invece per un altro che ha alle spalle una complessa storia criminale. Ciò dipende dal fatto che, come chiarito dalla Guida stessa, «*[r]isk assessment is about predicting group behavior (identifying groups of higher risk offenders) - it is not about prediction at the individual level. Your risk score is estimated based on known outcomes of groups of offenders who have similar characteristics*». Non è dunque

– e nemmeno potrebbe essere – una valutazione del rischio di recidiva individuale quanto piuttosto una valutazione del rischio del gruppo al quale il soggetto è stato identificato come appartenente³¹.

In questo caso, specifica la Guida, spetterà al giudice considerare adeguatamente le circostanze del caso e sovvertire (*override*) il risultato di COMPAS. Si legge nella Guida: «*we would expect staff to disagree with an actuarial risk assessment (e.g. COMPAS) in about 10% of the cases due to mitigating or aggravating circumstances which the computer is not sensitive to. In those cases staff should be encouraged to use their professional judgment and override the computed risk as appropriate - documenting it in COMPAS with the Override Reason - for monitoring by supervisory staff*»³².

²⁷ Occorre rimarcare il fatto che i fattori criminogeni utilizzati sono il frutto di una scelta della società produttrice del *software*, che ha accolto alcune teorie criminologiche che tuttavia non hanno unanime riconoscimento. In argomento si v. K. B. FORREST, *When machines can be judge, jury, and executioner: justice in the age of artificial intelligence*, cit., pp. 69-70.

²⁸ È possibile consultare un esempio di questionario COMPAS al seguente [link: Questionario COMPAS](#).

²⁹ Guida COMPAS, pp 31 ss.

³⁰ Guida COMPAS, p. 13, in cui si legge: «*Machine learning methods are highly flexible and are usually preferred in applications where there is a complex relation- ship between predictors and outcomes*».

³¹ Vedremo che una delle contestazioni che verranno mosse ai *tools* sarà proprio la mancanza di una valutazione di rischio individualizzata.

³² Guida COMPAS, p. 35.

Ora, attraverso i dati relativi ai fattori di rischio l'algoritmo di COMPAS, mediante procedure di *pattern recognition*³³, classifica automaticamente ogni individuo in una "tipologia"³⁴ che meglio si adatta al profilo del soggetto analizzato.

Le tipologie di COMPAS sono otto per gli uomini³⁵ e otto per le donne³⁶. Gli operatori, sulla base della tipologia standard assegnata all'individuo, basano il proprio *case plan* (e il giudice penale la determinazione dell'ammontare della pena).

Ultima caratteristica di COMPAS da mettere in luce per comprendere come funziona l'algoritmo al suo interno è il *dataset* utilizzato per la progettazione e l'addestramento (*training*) dell'algoritmo. In che modo, cioè, il sistema, elaborando gli *input* che provengono dai dati relativi a ciascun soggetto, individua una tipologia *standard* da assegnare all'imputato e il livello di rischio di recidiva. È sicuramente indicativo che solo poche righe della Guida siano dedicate al *dataset*³⁷, che è ciò che nutre l'algoritmo e ne determina il funzionamento: come si vedrà, uno dei problemi che pongono siffatti sistemi è l'accessibilità dei dati, e il disvelamento del *dataset* sarebbe già un primo passo per assicurare una maggiore trasparenza dell'algoritmo stesso.

Ad ogni buon conto, nella Guida si sottolinea che la "base di conoscenza" dell'algoritmo è costituita da un'indagine a campione su trenta mila persone svolta negli Stati Uniti tra il 2004 e 2005 su soggetti sia condannati e in carcere, sia non ancora condannati, sia in fase di esecuzione della pena al di fuori della struttura carceraria. All'interno del primo campione i progettisti, con modalità non rinvenibili e dunque non scrutinabili, hanno selezionato un campione ridotto di 7381 soggetti, che rappresentano il *dataset* di COMPAS; pertanto, le inferenze che compie l'algoritmo si fondano sostanzialmente sulla caratterizzazione criminale di tale gruppo campione. Benché la Guida non fornisca altre spiegazioni, i limiti dell'utilizzo di una simile base conoscitiva sono evidenti. In primo luogo, i dati si riferiscono a soggetti che erano nel circuito criminale tra il 2004 e il 2005 e non c'è stato alcun aggiornamento, di tal

³³ Il *pattern recognition* è una tecnica di *machine learning* che consiste nell'identificazione di modelli partendo da dati grezzi, al fine di identificare la corretta classificazione per i nuovi dati. Per un approfondimento si cfr. la pagina web [Introduction to Pattern Recognition in Machine Learning](#).

³⁴ Guida COMPAS, p. 51.

³⁵ In particolare, le otto tipologie in cui si può venire classificati sono: 1. "Chronic drug abusers – most non-violent"; 2. "Low risk situational – fighting/ domestic violence caution"; 3. "Chronic alcohol problems – DUI, domestic violence"; 4. "Socially marginalized – poor, uneducated, stressed, habitual offenders"; 5. "Criminally versatile – young marginalized persons often gang affiliated"; 6. "Socially isolated long term substance abuse – multiple minor and mostly non-violent offenses"; 7. "Serious versatile high risk individuals"; 8. "Low risk situational accidental category". Si v. Guida COMPAS, pp. 54 ss.

³⁶ In particolare: 1. "Drug problems and anti-social sub-cultural influences – some with relationship conflicts"; 2. "Family disorganization and inadequate parenting – residential in- stability and minor non-violent offences"; 3. "Chronic substance abusers – women with higher social resources than other groups"; 4. "Marginalized poor and isolated older women – economic survival crimes"; 5. "Young antisocial poorly educated women with some violent offences and early delinquency onset"; 6. "Chronic long term criminal history A – multiple co-occurring social and psychological risk factors"; 7. "Chronic long term criminal history B – multiple co-occurring problems and high risk"; 8. "Late starters with multiple strengths and fewer risk factors – minor non-violent offence history". Si v. Guida COMPAS, pp. 59 ss.

³⁷ Guida COMPAS, p. 11.

che COMPAS determinerà il rischio di recidiva di un individuo basandosi sui livelli di recidiva esistenti nei primi anni 2000³⁸. Inoltre, non sembra ci sia una differenziazione territoriale tra i dati raccolti (contestazione sollevata dinanzi alle Corti³⁹), onde una rappresentazione distorta di quelli che possono essere i diversi indici di recidiva dei diversi Stati o delle diverse regioni.

I problemi dell'algoritmo COMPAS che più sono stati oggetto di studio, discendenti dal suo funzionamento, sono l'effetto discriminatorio che esso *necessariamente*⁴⁰ produce e la sua *accuracy* predittiva⁴¹.

2.3 L'acquisizione del risultato algoritmico da parte del giudice nella fase di determinazione in concreto della pena

Chiarito il funzionamento dei *tools* ai fini che qui interessano, per meglio comprendere come la giurisprudenza americana, peraltro in modo non sempre convincente, ha tentato di sciogliere i nodi connessi al rapporto tra giudice e strumenti algoritmici nella fase del *sentencing*⁴², occorre preliminarmente chiarire in quali termini, in sede processuale, il giudice acquisisce le risultanze del sistema algoritmico.

Una volta che il reo viene dichiarato colpevole, di norma il giudice richiede il *Pre-sentence Investigation Report* (PSI o PSIR), un rapporto predisposto solitamente da un funzionario della Corte contenente dati e

³⁸ Si pensi, ad esempio, a una giurisdizione statale che utilizza COMPAS per fare una predizione sull'attuale rischio di un imputato afroamericano, utilizzando dati di un periodo in cui (nella specie all'inizio degli anni 2000) l'arresto per consumo di marijuana era particolarmente elevato per le minoranze etniche. Si cfr. K. B. FORREST, *When machines can be judge, jury, and executioner: justice in the age of artificial intelligence*, cit., p. 77.

³⁹ Si v. *infra*, § 5.

⁴⁰ È facilmente intuibile (anche dall'esempio indicato alla nota 38), che quando i dati sono raccolti all'interno di un sistema giurisdizionale discriminatorio, l'algoritmo riprodurrà le medesime discriminazioni di quel sistema (in informatica si usa l'efficace espressione *garbage in, garbage out*). Per un approfondimento si cfr. C. ORWAT, *Risks of Discrimination through the Use of Algorithms*, in *Federal Anti-Discrimination Agency*, 2020 e S. A. FRIEDLER - C. SCHEIDEGGER - S. VENKATASUBRAMANIAN, *On the (im)possibility of fairness*, in *ArXiv*, 23 settembre 2016.

⁴¹ Come è evidente, si tratta di problematiche che, sebbene assai interessanti, solo lateralmente toccano il tema indagato nel presente contributo e che quindi non saranno oggetto di specifico approfondimento. Sulla tendenza dei *tools* a riprodurre discriminazioni esistenti nonché sui difetti di affidabilità, si cfr., *ex multis*, J. KLEINBERG - J. LUDWIG - S. MULLAINATHANY - C. R. SUNSTEIN, *Discrimination in the age of algorithms*, in *Journal of Legal Analysis*, vol. 10, 2018, pp. 113-174; J. ANGWIN - J. LARSON - S. MATTU - L. KIRCHNER, *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, *ProPublica*, 23 maggio 2016; S. GOEL - R. SHROFF - J. L. SKEEM - CHRISTOPHER SLOBOGIN, *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment*, cit.; A. Z. HUQ, *Racial equity in algorithmic criminal justice*, in *Duke Law Journal*, vol. 68 (6), 2019, pp. 1043-1134; T. Z. ZARSKY, *An Analytic Challenge: Discrimination Theory in the Age of Predictive Analytics*, in *A Journal of Law and Policy for the Information Society*, vol. 14 (1), 2017, pp. 11-35; J. VILLASENOR - V. FOGGO, *Artificial intelligence, due process, and criminal sentencing*, cit.; M. BRENNEN *et al.*, *Constitutional dimensions of predictive algorithms in criminal justice*, cit.; C. WINTER, *The challenges of artificial judicial decision-making for liberal democracy*, Working Paper n. 3 del 2021.

⁴² Meno problemi pone l'impiego degli strumenti predittivi nella fase pre-processuale (*pre-trial*) dove l'obiettivo della valutazione del rischio è relativamente ben definita; il giudice deve prevedere se l'imputato si presenterà in tribunale e la probabilità che, nel frattempo, commetta qualche reato. Se l'imputato presenta un rischio di fuga o un pericolo per il pubblico, il giudice probabilmente scongiurerà il rilascio, e viceversa. Non sorprende, quindi, che sia diventato sempre più frequente integrare il processo con un *software* di valutazione del rischio, per contribuire a ridurre il numero di persone in carcere prima della celebrazione del processo. Si cfr. D. KEHL - P. GUO - S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, cit., p. 13.

informazioni inerenti al soggetto condannato, che il giudice è chiamato a valutare esercitando il massimo grado di discrezionalità (*great o board discretion*⁴³). Al riguardo, nella fase di *sentencing* i limiti normativi rispetto alla valutazione discrezionale delle prove e degli altri elementi che il giudice compie prima della determinazione finale sulla pena da applicare sono estremamente esigui: se la fase più propriamente processuale, e cioè la parte di *trial* il cui esito conclusivo sfocia nella dichiarazione di colpevolezza (*plea of guilty*), è informato al rispetto di rigidi confini nella valutazione delle prove, nella fase di determinazione della pena il giudice non è rigidamente vincolato ai fatti rilevanti nel giudizio, ma la sua valutazione deve considerare, nell'ottica di reinserimento del condannato nella società, anche elementi specifici della personalità dello stesso⁴⁴.

Negli Stati che hanno, esplicitamente o meno⁴⁵, adottato negli ultimi anni i *tools* di predizione del rischio di recidiva, all'interno del PSI vengono riportati anche i risultati dell'indagine realizzata attraverso il sistema algoritmico⁴⁶, il quale indica, come visto, un indice di rischio di recidiva che varia tra alto (*high*), medio (*medium*) e basso (*low*) di cui il giudice potrà tenere conto per determinare le misure trattamentali più idonee per il condannato e se possa scontare la pena al di fuori delle strutture carcerarie⁴⁷.

3. L'ampia casistica giurisprudenziale sui *risk assessment tools*: non solo *Loomis*

La breve descrizione del funzionamento dei *tools* e di come essi sono impiegati in sede processuale è funzionale alla disamina critica della giurisprudenza che è maturata in seno agli Stati Uniti rispetto all'utilizzo di siffatti strumenti.

Lo sguardo prospettico adottato in questa sede non mira a sviluppare ulteriormente le preoccupazioni, da tempo manifestate, rispetto alle garanzie fondamentali del giusto processo; piuttosto, l'obiettivo è guardare il fenomeno con gli occhi del giudice penale, per comprendere come muta l'esercizio del potere discrezionale al cospetto di un algoritmo che contribuisce a formare la sua base di conoscenza ai fini del decidere.

⁴³ La giurisprudenza, sia a livello federale sia a livello statale, è assolutamente costante. Si v., di recente, in ambito federale, SUPREME COURT OF THE UNITED STATES, *Concepcion v. United States*, 27 giugno 2022.

⁴⁴ Per un approfondimento si v., ancora, D. KEHL - P. GUO - S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, cit., pp. 15 ss.

⁴⁵ La "base legale" che autorizza l'impiego dei sistemi predittivi varia da Stato a Stato. Mentre alcuni Stati richiedono esplicitamente che venga inserito obbligatoriamente nel PSI anche l'algoritmo predittivo ai fini dei *criminal sentencing* (per esempio Arizona e Oklahoma), altri si limitano a permetterne l'utilizzo senza però imporlo (per esempio Idaho); si cfr. KEHL - P. GUO - S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, cit., pp. 15 ss.

⁴⁶ Al sito di NATIONAL CENTER OF STATE COURTS, che da più di cinquant'anni raccoglie dati delle Corti dei diversi Stati, è possibile visionare un esempio di una sezione del PSI contenente i risultati del *risk assessment tool*: si rinvia quindi alla relativa [pagina web](#).

⁴⁷ Per uno studio su come i giudici interpretano i risultati nei *risk assessment tools*, ancorché nei casi in cui essi sono utilizzati nel *pre-trial*, si cfr. M. DEMICHELE - M. COMFORT - K. BARRICK - P. BAUMGARTNER, *The Intuitive-Override Model: Nudging Judges Toward Pretrial Risk Assessment Instruments*, in *Federal Probation Journal*, vol. 85, n. 2, 2021.

Il perseguimento di detto obiettivo non può che muovere, quindi, dalla giurisprudenza che ha scrutinato il nuovo perimetro del potere discrezionale di fronte all'algoritmo predittivo.

Quando si ragiona della giurisprudenza americana sui *risk assessment tools*, il primo riferimento è costituito dal celebre caso *Loomis*, la cui risonanza avuta sui *media* americani⁴⁸, non di rado frutto di fraintendimenti, ha rappresentato la prima occasione nella quale si è discusso di giudici *robot* e, in generale, delle opportunità e dei pericoli connessi al ricorso all'intelligenza artificiale nelle dinamiche processuali⁴⁹.

Il caso *Loomis*, lungi dall'essere isolato, si inserisce invero in un orientamento giurisprudenziale piuttosto ampio e variegato, ad esso antecedente e posteriore, sinora rimasto pressoché inesplorato e utile a disvelare come le Corti statali americane hanno vagliato la legittimità dei *risk assessment tools* nel *sentencing*, in particolare in procedimenti in cui veniva contestato dalle parti un abuso di discrezionalità del giudice (*abuse of discretion*).

Mentre a livello federale ancora non si è presentata l'occasione per esaminare la questione⁵⁰, nelle giurisdizioni statali – in particolare l'Indiana, il Wisconsin, il Michigan e lo Iowa – si è sviluppato un numero rilevante di pronunce da cui è possibile trarre alcune tendenze che appaiono in costante evoluzione.

4. (Segue) Critiche sistematiche e ordinamentali all'utilizzo nel *sentencing* degli strumenti predittivi

Nelle prime pronunce rinvenibili sull'impiego dei *tools* nel *sentencing*, adottate in Indiana, prevale indubbiamente forte cautela.

⁴⁸ Si cfr. ad esempio, il titolo sensazionalistico di una delle prime notizie apparse sul più importante giornale americano rispetto al caso *Loomis*: A. LIPTAK, *Sent to Prison by a Software Program's Secret Algorithms*, The New York Times, maggio 2017.

⁴⁹ Con riguardo alla letteratura straniera, si cfr., in particolare, l'accurata ricerca di T. SOURDIN, *Judges, Technology and Artificial Intelligence. The artificial Judge*, Cheltenham (UK), 2021 e, prima, T. SOURDIN, *Judge v Robot? Artificial intelligence and judicial decision-making*, in *UNSW Law Journal*, 41, 2018, pp. 1128 ss.. Si cfr. anche J. ULENAERS, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, cit.. In Italia la letteratura costituzionalistica che si è occupata dell'impiego dell'intelligenza artificiale nei sistemi giudiziari è consistente. Si segnalano tra i tanti, per l'acutezza delle riflessioni, M. LUCIANI, *La decisione giudiziaria robotica*, in A. CARLEO, *Decisione robotica*, Il Mulino, Bologna, 2019; F. DONATI, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1/2020, pp. 415 ss.; C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE online*, n. 3, 2020, pp. 3369 ss.; C. SALAZAR, *Judex ex Machina? Note su giustizia, giudici e intelligenza artificiale*, in *ConsultaOnline*, 18 novembre 2021; A. MORELLI, *Il giudice robot e il legislatore naif: la problematica applicazione delle nuove tecnologie all'esercizio delle funzioni pubbliche*, in *ConsultaOnline*, 6 agosto 2020.

⁵⁰ Nel caso *Loomis*, a seguito della decisione del rigetto da parte Suprema Corte del Wisconsin (si v. *infra*, § 5), l'imputato aveva richiesto il *Writ of Certiorari* alla Corte suprema degli Stati Uniti, uno strumento con cui si chiede una *judicial review* di una sentenza di una Corte statale. Nel giugno del 2017 la *US Supreme Court* negò il *certiorari*, ma è comunque interessante segnalare che nel [documento depositato](#) per opporsi alla richiesta di revisione da parte di *Loomis*, il Governo statunitense riconosce che «*some uses of an undisclosed risk-assessment algorithm might raise due process concerns if, for example, a defendant is denied access to the factual inputs out his criminal and personal history, or if his risk scores form part of a sentencing 'matrix' or establish a 'presumptive' term of imprisonment*».

Nello specifico, già nel 2008 due sentenze risultano particolarmente significative in quanto toccano, seppur marginalmente, l'impatto degli algoritmi sul processo discrezionale del giudice. Nella pronuncia *Cardwell v. State*⁵¹, invero, non veniva in gioco alcuna applicazione algoritmica per la determinazione del rischio di recidiva: ciononostante, all'esito di un procedimento promosso dinanzi alla Corte suprema per *abuso di discrezionalità* del giudice circondariale, in quanto questi non aveva considerato alcune circostanze attenuanti, la Corte suprema dell'Indiana coglie l'occasione per svolgere alcune considerazioni di portata generale sull'impiego di "automatismi tecnologici" nell'ambito del *sentencing*.

In un passo della sentenza la Corte suprema afferma che il sistema giurisdizionale dell'Indiana «*has never adopted a mechanical approach to sentencing, and we have not identified any inflexible system that did not raise more problems than it solved*». In modo ancor più netto la Corte ritiene che «*[a]ny effort to force a sentence to result from some algorithm based on the number and definition of crimes and various consequences removes the ability of the trial judge to ameliorate the inevitable unfairness a mindless formula sometimes produces*». Secondo la Corte, quindi, qualunque sforzo teso a far "dipendere" una sentenza dall'algoritmo impedisce al giudice di correggere l'«inevitabile iniquità» che talvolta un procedimento matematico «*mindless*» produce.

Le parole della *Supreme Court* vengono riprese nello stesso anno dalla sentenza *Rhodes v. State*⁵² pronunciata dalla Corte d'Appello dell'Indiana che, sviluppando in concreto le considerazioni generali svolte in *Cardwell* in un caso in cui vi era stato l'effettivo utilizzo da parte del giudice di uno strumento di *risk assessment* (LSI-R, di terza generazione)⁵³ al fine di determinare l'ammontare della pena, accoglie l'appello riscontrando un abuso di discrezionalità da parte della Corte distrettuale in quanto l'utilizzo di siffatti *tools* da parte del tribunale per commisurare la severità della pena «*is contrary to the essential function of the trial court in sentencing*». Secondo la Corte d'appello il ruolo che è assegnato al giudice nel *sentencing* gli impone, infatti, di prendere in considerazione *autonomamente* i fattori contenuti nel PSI (le circostanze del fatto, la storia criminale dell'imputato, la condizione familiare, lavorativa, sociale e così via), il cui compito è quello di fornire al giudice le informazioni necessarie per «*create an appropriate, individualized sentence*». In altre parole, i fattori che in un sistema algoritmico rappresentano i dati di *input* sui quali la macchina costruisce la sua "previsione" devono essere valutati *autonomamente* dal giudice nel libero esercizio del suo potere discrezionale.

⁵¹ *Cardwell v. State*, 895 N.E.2d 1219 (2008). Il caso riguardava la condanna di un uomo a trentaquattro anni di carcere per lesioni dolose a danni di una minore. Il ricorrente aveva impugnato la sentenza per *abuse of discretion* per non avere la Corte territoriale valutato attentamente tutte le circostanze attenuanti.

⁵² *Rhodes v. State*, 896 N.E.2d 1193 (2008). Nel caso di specie, due uomini, tra cui il Sig. Rhodes, venivano fermati dalla polizia alla guida di un'auto rubata e al cui interno venivano trovate dosi massicce di droga. A seguito di patteggiamento, il Sig. Rhodes veniva condannato a trent'anni di carcere per il reato di spaccio e appellava la sentenza per abuso di discrezionalità e, in particolare, per aver il giudice considerato due *improper aggravating factors* tra cui lo *score* prodotto dall'algoritmo.

⁵³ Si v. *supra*, § 2.2.

Solo così, secondo il giudice dell'Indiana, si evita una deresponsabilizzazione del giudice nella determinazione della pena appropriata, in quanto «*[t]he use of a standardized scoring model, such as the LSI-R, undercuts the trial court's responsibility to craft an appropriate, individualized sentence.* E ancora, affidarsi a una serie di numeri apparentemente derivati da dati oggettivi «*cannot serve as a substitute for an independent and thoughtful evaluation of the evidence presented for consideration*».

Da siffatte affermazioni traspare la preoccupazione di un inevitabile inaridimento della discrezionalità del giudice nel valutare gli elementi rilevanti per la decisione, con il conseguente venire meno della responsabilità dello stesso nel formulare una sentenza individualizzata. Su tali argomenti, come si vedrà, si appuntano le critiche dei successivi arresti giurisprudenziali che, in esplicita ed espressa contrapposizione a *Cardwell* e *Rhodes*, indirizzeranno il dibattito sulla legittimità dell'impiego dei *tools* nel *sentencing*.

5. (Segue) I primi argomenti a favore della legittimazione dell'uso dei *tools* da parte del giudice

Le voci contrarie alle profonde critiche mosse in *Cardwell* e *Rhodes* si sollevano sempre in Indiana, dove due sentenze coeve del 2010 della *Supreme Court*, pur nelle differenze della fattispecie oggetto del giudizio, negano la configurabilità di un abuso di discrezionalità da parte della Corte distrettuale, e adottano una posizione assai permissiva rispetto all'uso dei *risk assessment tools*.

In *Malenchik v State*⁵⁴ dapprima i giudici supremi precisano – ed è una precisazione significativa che verrà ripresa nelle pronunce successive – che la decisione sulla determinazione della pena era basata non solo sulle risultanze dei due sistemi predittivi utilizzati ma su altri, autonomi, fattori: la Corte distrettuale pertanto «*did not rely on either the LSI-R or SASSI as an independent aggravating factor in deciding to impose more than the advisory sentences*»⁵⁵.

Rispetto alle critiche rivolte nella sentenza *Cardwell* all'impiego di pratiche automatizzate nel *sentencing* (si cfr. *supra* al precedente §), la Corte se ne smarca rapidamente ponendo l'accento sull'obiettivo, insito nella giustizia penale, che «*an individualized penal consequence should be efficacious in achieving the goals of reformation and minimizing recidivism*» e per tale obiettivo l'approccio *evidence-based* dei *tools* «*has considerable promise*». Nel trade-

⁵⁴ *Malenchik v. State*, 928 N.E.2d 564, 571–73, 2010. Nel caso di specie, il soggetto era stato imputato per il reato di ricettazione e condannato a sei anni di carcere dopo che il giudice, tramite l'impiego di due sistemi algoritmici (LSI-R e SASSI), aveva considerato il Sig. Malenchik un *habitual offender*. Anche in questo caso la sentenza veniva impugnata per abuso di discrezionalità in quanto secondo il condannato la Corte avrebbe fatto affidamento sullo *score* prodotto dai due algoritmi.

⁵⁵ In questo caso il soggetto condannato sosteneva, sulla scorta del caso *Rhodes*, che erroneamente la Corte aveva considerato l'esito (*score*) del sistema algoritmico come circostanza aggravante nella determinazione della pena nonché lamentava la mancanza di «*objective reliability*» oltre che il fatto che tali sistemi erano «*unfairly discriminatory*».

off⁵⁶ tra giustizia del processo ed efficienza della macchina giudiziaria sembra qui prevalere nettamente la seconda.

Più specificamente, la *Supreme Court*, confrontandosi con le motivazioni fornite dalla Corte d'Appello in *Rhodes* relative alla deresponsabilizzazione del giudice nel momento in cui questi è chiamato a decidere con l'ausilio di modelli di valutazione standardizzati quali i *tools*, afferma perentoriamente: «*[w]e disagree*». Secondo la Corte, all'opposto, il punteggio prodotto dall'algoritmo consente alla valutazione individualizzata di rinvigorirsi proprio grazie al «*sound discernment and discretion*» con cui il giudice dovrebbe assegnare agli strumenti predittivi «*proper consideration and appropriate weight*». Detti strumenti, secondo la Corte, non sono infatti pensati per sostituire la decisione sulla gravità della pena, ma semmai per adiuvarne il decidente in valutazioni collaterali quali la scelta di un programma di libertà vigilata o programmi di trattamento alternativi alla detenzione; in altre parole, i punteggi che determinano il rischio di recidiva non sono necessariamente congruenti all'autonoma valutazione del giudice di tutti gli elementi rilevanti nel giudizio⁵⁷.

In *J.S. v. State*⁵⁸, coeva a *Malenchik*, viene in risalto il medesimo profilo visto da una prospettiva opposta: la contestazione di un abuso di discrezionalità non per aver la Corte fatto proprie le risultanze dell'algoritmo, ma per averle completamente trascurate.

Contrariamente a quanto visto in *Malenchik*, nel caso di specie il giudice distrettuale disattende il punteggio prodotto dall'algoritmo, che indicava un indice di recidiva molto basso, e applica una pena assai elevata⁵⁹. Mutuando le motivazioni spese in *Malenchik*, la Corte ha buon gioco nel rigettare l'abuso di discrezionalità in quanto il sistema algoritmico non era da considerarsi alla stregua di una circostanza aggravante o attenuante – gli strumenti di valutazione non sono infatti preparati dal giudice che emette la sentenza, ma dai funzionari dei dipartimenti, e le loro valutazioni possono non coincidere con le conclusioni del giudice del processo – quanto piuttosto solo uno strumento per «*supplement and enhance a judge's evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant*».

In definitiva, la giurisprudenza dell'Indiana assume un chiaro orientamento: non c'è abuso di discrezionalità da parte del giudice penale né quando questi valorizza il risultato dell'algoritmo predittivo,

⁵⁶ M. SUN - M. GERCHICK, *The Scales of (Algorithmic) Justice: Tradeoffs and Remedies*, in *Ai Matters*, vol. 5 (2), 2019, pp. 30 ss.

⁵⁷ In dottrina, sugli altri profili suggeriti dal caso *Malenchik*, si cfr. R. BERK - J. HYATT, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, in *Federal Sentencing Reporter*, 27, no. 4, 2015 pp. 226 ss.; J. VILLASENOR - V. FOGGO, *Artificial intelligence, due process, and criminal sentencing*, cit., spec. p. 337.

⁵⁸ *J.S. v. State*, 928 N.E.2d 576 (2010). Nel caso di specie l'imputato veniva dichiarato colpevole per due distinti capi di accusa concernenti molestie su minori e gli veniva comminata una pena di otto anni per il primo e di quattro per il secondo.

⁵⁹ Si legge nell'estratto della motivazione del Giudice distrettuale: «*[i]m going to discount the LSI-R. I've been advised that that's not a good measure in cases involving sexual abuse, so I'm not going to treat that either as an aggravator or a mitigating factor. I do find that the aggravating factors outweigh the mitigating factors*».

né tantomeno quando se ne discosta, essendo quest'ultimo solo uno dei tanti elementi che il decidente deve considerare per assumere la propria decisione.

Su posizioni simili si assesta la giurisprudenza maturata in seno allo Stato del Wisconsin, particolarmente significativa e non limitata al solo, affascinante, caso *Loomis*, ma estesa, come detto, ad altri casi ad esso precedenti e successivi.

In *Samsa*⁶⁰, l'abuso di discrezionalità lamentato concerneva, come in *J.S.* (si v. *supra*), l'interpretazione del risultato di COMPAS, il quale segnalava alcuni fattori criminogeni (i *criminogenic needs* di cui abbiamo fatto cenno⁶¹) dell'imputato, sulla base dei quali tuttavia produceva un indice di rischio di recidiva basso (*low risk*). La Corte distrettuale non condivideva l'esito finale dello *score* algoritmico e, attraverso una valutazione autonoma dei fattori criminogeni indicati da COMPAS, condannava Samsa a una pena elevata. Di fronte alla contestazione del reo, il quale lamentava un *abuse of discretion* per non essersi il giudice attenuto al risultato dell'algoritmo, la Corte d'appello del Wisconsin precisa che il *sentencing* deve rimanere oggetto di «*great discretion*» e che il rifiuto considerare l'esito algoritmico «*was perfectly within the court's discretion*», in quanto «*[t]he COMPAS is merely one tool available to a court at the time of sentencing and a court is free to rely on portions of the assessment while rejecting other portions*».

Due anni dopo *Samsa* giunge all'attenzione della Corte il celebre caso *Loomis*⁶², che, come già si è anticipato, ha avuto il merito di portare il dibattito su un piano diverso, in quanto per la prima volta entrava in gioco in modo esplicito il legame esistente tra la previsione algoritmica del rischio di recidiva e i principi del giusto processo⁶³.

Qui, tuttavia, non ci si sofferma sulle numerose problematiche connesse al *due process* rispetto alle quali la dottrina si è concentrata, quanto piuttosto su come gli elementi di contestazione sollevati dall'imputato – nella specie, in particolare, l'impossibilità di accedere al codice dell'algoritmo stante la natura proprietaria del *software* e la mancata individualizzazione della sentenza stante le caratteristiche di COMPAS – incidono sulla riconfigurazione dell'esercizio della discrezionalità del giudice.

⁶⁰ *State v. Samsa*, 359 Wis.2d 580 (2014). Nel caso in esame, il Sig. Samsa veniva condannato alla pena di cinque anni di carceri più cinque di *extended supervision* per il reato di violenza sessuale, nonostante nel PSI venisse raccomandato, sulla base dell'algoritmo COMPAS, la libertà vigilata per venire incontro ai *criminogenic needs* del soggetto condannato.

⁶¹ Si v. *supra*, § 2.2.

⁶² *State v. Loomis*, 881 N.W.2d 749 (Wis 2016). Nel caso in esame Loomis veniva accusato dal Wisconsin sulla base di cinque distinti capi d'imputazione in relazione a una sparatoria in auto e, a seguito di una valutazione di rischio effettuata da COMPAS e inserita nel PSI, veniva condannato a sei anni di carcere e cinque di *extended supervision*.

⁶³ In particolare, l'imputato sollevava tre distinte doglianze afferenti alla violazione del *due process* perpetrate dall'utilizzo di COMPAS nel *sentencing*: la natura proprietaria di COMPAS impediva di accedere all'algoritmo e valutarne l'attendibilità (*accuracy*) in quanto non era possibile verificare come viene determinato l'*output* del sistema e il peso (*weight*) di ogni singolo fattore considerato; l'affidamento da parte del giudice su un sistema automatico di valutazione del rischio violava il suo diritto a una *individualized sentence*; essendo presente, nei dati elaborati da COMPAS, anche il genere come fattore criminogeno, la considerazione della valutazione del rischio da parte della Corte circondariale violava il diritto costituzionale a non essere giudicato sulla base del genere.

Rispetto a quest'ultimo aspetto, la conclusione cui perviene la Corte suprema del Wisconsin, in armonia con quanto già affermato dalla Suprema Corte dell'Indiana in *Malenchik*, è che la Corte circondariale non abbia abusato della propria discrezionalità, in quanto la considerazione da parte del giudice del punteggio di COMPAS non è stato un fattore decisivo nella scelta⁶⁴ e, di conseguenza, la Corte avrebbe usato COMPAS solo per corroborare una decisione sulla pena che sarebbe stata determinata nel medesimo ammontare anche senza l'algoritmo⁶⁵.

Detta conclusione viene raggiunta dopo una lunga e meditata motivazione indirizzata a respingere l'argomento di *Loomis* secondo il quale l'adesione al risultato algoritmico, basato su dati statistici appartenenti a un determinato gruppo (come si è visto discutendo delle caratteristiche dei *tools*, si v. *supra*, § 2.2) priverebbe l'imputato del suo diritto a ottenere una sentenza individualizzata, la quale dovrebbe essere parametrata non su risultati automatizzati ma decisa discrezionalmente dal giudice.

Secondo la Corte, al contrario, non essendo COMPAS l'unico elemento su cui si fonda la decisione, «*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*». In sostanza, i giudici affermano: «*we expect that circuit courts will exercise discretion when assessing a COMPAS risk score with respect to each individual defendant*»⁶⁶.

Più di recente, anche in *Michigan*, dove da alcuni anni le valutazioni di COMPAS sono entrate a far parte del PSI⁶⁷, la Corte d'appello perviene a conclusioni simili a quelle espresse dalla giurisprudenza sopra citata.

In *People v. Younglove*⁶⁸ è ancora l'argomento della non esclusività della prova algoritmica a costituire la *ratio decidendi* per rigettare le doglianze afferenti alle violazioni del giusto processo. Nella specie, alle contestazioni afferenti all'assenza di una sentenza individualizzata in quanto l'analisi algoritmica concerneva dati appartenenti all'intera popolazione (si v., analogamente, *Loomis*), i giudici di appello rispondono che «*[a] sentencing court is not bound by the recommendations included in a PSIR [...]. Nor is a sentencing court obliged to use all information contained within a PSIR in pronouncing sentence. Courts must consider the PSIR, but whether or how heavily to weigh the information contained therein remains within the court's discretion*».

⁶⁴ *Ivi*, §§ 8 e 9.

⁶⁵ *Ivi*, § 28.

⁶⁶ *Ivi*, § 72.

⁶⁷ Si cfr. la esplicativa guida prodotta dal *Department of Correction* dello Stato del Michigan al presente [link](#).

⁶⁸ *People v. Younglove*, No. 341901, 2019. Nel caso in esame i giudici di appello prendevano in considerazione i gravami di diversi imputati le cui condanne erano state determinate anche con l'utilizzo del sistema COMPAS: il Sig. Younglove ricorreva in appello contro la condanna a un anno di carcere e tre anni di libertà vigilata inflitta dal tribunale per il possesso di pseudoefedrina per la produzione di metanfetamina; il Sig. Hegler appellava la condanna da 1 anno e mezzo a 5 anni di reclusione che il tribunale comminava per guida di veicolo in stato di ebbrezza; il Sig. Wilson impugnava la decisione (dai venti ai sessanta mesi di reclusione) per frode; il Sig. Bradford ricorreva in appello contro la sentenza da 1 anno e mezzo a 5 anni di reclusione per furto di arma da fuoco.

In definitiva, affermano i giudici, «*we are not persuaded by defendants' arguments that the inclusion of COMPAS information unfairly influences or replaces a sentencing court's individual sentencing discretion*».

6. (Segue) Il peso dell'algoritmo nel percorso discrezionale del giudice: il labile confine tra *reliance* e *consideration*

Dalle sentenze analizzate è possibile individuare un filo rosso che intreccia e unisce le motivazioni dei giudici: il peso e la rilevanza l'algoritmo assume nella valutazione discrezionale del giudice sul *quantum* della pena.

In *Cardwell* e *Rhodes* l'algoritmo viene dipinto come uno strumento «inevitabilmente ingiusto» in quanto privo di capacità discrezionale; l'utilizzo del risultato algoritmico contrasta con la intima funzione del giudice penale che, discrezionalmente, deve poter valutare i fattori di pericolosità del soggetto sottoposto al proprio giudizio sulla cui base formulare una sentenza appropriata e individualizzata.

Nel giro di pochi anni, invece, la giurisprudenza che matura all'interno dell'Indiana e del Wisconsin, seguita di recente dalla giurisprudenza del Michigan, inizia a distinguere tra *reliance* e *consideration* del risultato standardizzato prodotto dall'algoritmo decisionale: se vi è unanime giudizio sull'impossibilità per il giudice di appiattirsi (*rely on*) esclusivamente sul risultato algoritmico quale unico fattore rilevante per la decisione, quest'ultimo diviene uno degli elementi che il giudice ha la facoltà sia di considerare per addivenire a una sentenza individualizzata (si cfr. *Malenchik*, *Loomis* e *People v. Younglove*), sia di trascurare e dare peso a elementi ulteriori e diversi (si cfr. *Samsa* e *J.S.*).

La delicata distinzione tra *reliance* e *consideration*, se per un verso può apparire chiara, suscita molteplici incertezze.

La confusione sembra emergere in modo cristallino in una *concurring opinion* della Giudice Roggensack nel caso *Loomis*, in cui la stessa, pur condividendo la decisione della maggioranza rispetto alla legittimità dell'utilizzo dei *tools*, evidenzia che la Corte, in motivazione, ha erroneamente utilizzato in maniera intercambiabile i due termini (*reliance* e *consideration*), così rischiando di legittimare surrettiziamente anche decisioni che assumono le risultanze dell'algoritmo come fattore *decisivo* per la determinazione della pena⁶⁹. Nella valutazione discrezionale del giudice sul *quantum* della pena, risulta infatti arduo verificare quale peso sia stato effettivamente attribuito al risultato algoritmico, tanto più, come sovente accade, nel caso di algoritmi di *machine learning* che producono un indice di recidiva senza fornire una spiegazione del procedimento che ha condotto alla produzione del risultato.

⁶⁹ Queste le parole della Giudice Roggensack: «*[b]ecause at times the majority opinion interchangeably employs consider and rely when discussing a sentencing court's obligations and the COMPAS risk assessment tool, our decision could mistakenly be read as permitting reliance on COMPAS. Therefore, I write to clarify for the reader*», cfr. *Loomis*, § 124.

Quale che sia il risultato del sistema predittivo, se il giudice assume la propria decisione *considerando* il risultato dell’algoritmo e dunque conformando la sentenza a detto risultato, il confine con la *reliance* appare quanto mai labile, se non addirittura inesistente⁷⁰.

A conclusioni analoghe perviene anche una pronuncia dello Stato dell’Iowa nel 2018 – *State v. Gordon*⁷¹, in un caso in cui l’imputato contestava l’*abuse of discretion* della Corte distrettuale, che si sarebbe affidata integralmente (*rely*) al risultato algoritmico. Rispetto alla controdeduzione dello Stato – che invece eccepiva la sola *considerazione* dell’algoritmo, e quindi l’assenza di una vera e propria *reliance* – la Corte d’appello precisa: «*[w]e are not persuaded that the difference between reliance upon and consideration of these actuarial estimates saves the sentencing process in this case*».

Secondo la Corte, quindi, non è agevole individuare il peso che sia stato assegnato a un singolo fattore (la Corte dice, efficacemente, «*mentally assigned*») e se esso sia stato decisivo rispetto alla decisione sulla pena. In definitiva, la Corte d’appello dello Iowa rinviene un abuso di discrezionalità del giudice territoriale in quanto ritiene troppo sottile è il confine tra *reliance* e *consideration* del sistema algoritmico⁷².

Alla luce di tali approdi si possono svolgere alcune prime considerazioni di carattere generale.

Il rischio concreto che si corre con l’impiego di algoritmi predittivi in sede decisione è che un risultato automatizzato, frutto di un procedimento meccanico apparentemente scevro da condizionamenti esterni⁷³, percepito come oggettivo ed esatto, assuma, nella decisione del giudice, un peso preponderante e quasi esclusivo⁷⁴: ciò deriva sostanzialmente da quella che è stata definita “forza di argomentazione pratica” dell’algoritmo⁷⁵ ossia la tendenza dell’algoritmo ad attrarre la decisione stessa per ragioni di

⁷⁰ Tali preoccupazioni sono manifestate, sebbene con estrema sinteticità, anche in S. GOEL - R. SHROFF - J. L. SKEEM - C. SLOBOGIN, *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment*, dicembre 2018, disponibile in [SSRN](#), p. 17.

⁷¹ *State v. Gordon*, 919 N.W.2d 635, 2018. Nel caso di specie, Gordon veniva condannato a una condanna di non più di dieci anni per violenza sessuale nei confronti di una minore. Nel PSI erano contenuti i punteggi ottenuti da due strumenti di valutazione del rischio: lo STATIC-99R e la *Sex Offender Treatment Intervention and Progress Scale* (SOTIPS). Il punteggio dello STATIC-99R indicava che Gordon era di rischio medio di recidiva. Il punteggio SOTIPS indicava che era un individuo ad alto rischio di recidiva.

⁷² La sentenza della Corte d’appello verrà poi ribaltata dalla Corte suprema dello Iowa per ragione più processuali che sostanziali e che quindi qui non interessano. Si v. *State v. Gordon*, 921 N.W.2d 19, 2018. Tuttavia, alcune parole dei Supremi Giudici ci saranno utili per sviluppare le nostre argomentazioni in merito alla trasparenza algoritmica, si v. *infra*, § successivo.

⁷³ Sull’inganno della “neutralità” dell’algoritmo la letteratura è vastissima. Si v. tra i molti, A. SIMONCINI, *L’algoritmo incostituzionale; intelligenza artificiale e il futuro delle libertà*, in *Rivista di BioDiritto*, 1, 2019, p. 81; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, fascicolo speciale, maggio 2019, p. 121. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, cit., pp. 4-6.

⁷⁴ Cfr. C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE online*, 2020 (3), p. 3388; si v. anche, per quanto riguarda il medesimo problema nella giustizia civile, C. BARBARO - Y. MENECEUR, *Issues in the use of artificial intelligence (AI) algorithms in judicial systems*, in *European Commission for the Efficiency of Justice Newsletter*, Council of Europe, n. 16, agosto 2018 e C. BARBARO, *Uso dell’intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, in *Questione giustizia*, 4, 2018, p. 194.

⁷⁵ L’espressione compare in A. SIMONCINI - S. SUWEIS, *Il cambio di paradigma nell’intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1, 2019, p. 100. Già prima Simoncini parlava di «travolgente forza pratica

convenienza pratica⁷⁶. Una volta, infatti, che si ingaggia l'algoritmo come “prova principe” – caratteristica che possiede per sua stessa natura, come abbiamo spiegato – risulta molto agevole per il giudice assumere la propria decisione motivandola alla luce di quello che l'algoritmo gli ha indicato (sul ruolo della motivazione si v. *infra*, § 8). L’*“algorithmic authority”*, insomma, impedisce al giudice di poter vagliare attentamente tutti gli elementi presenti e finisce per attrarre la decisione finale, senza che questa conclusione venga scalfita dalla considerazione secondo cui il giudice avrebbe solo *considerato* l'algoritmo. In definitiva, l'abuso di discrezionalità, spesso non riscontrato dai giudici che si sono rifugiati nella evanescente distinzione tra *decisività* e *complementarità* dell'algoritmo, disvela il mutamento dell'esercizio discrezionale del potere dinanzi ai sistemi di predizione del rischio e trova un suo primo punto di ricaduta nella impossibilità quasi immanente di distinguere tra un obbligo – affidarsi integralmente a un risultato automatizzato – e, all'opposto, una mera facoltà di attenersi.

Tale parziale conclusione conduce pertanto a riflettere della legittimità intrinseca dell'utilizzo dei *tools*, considerato che questi assorbono la discrezionalità del giudicante la quale, di conseguenza, si inaridisce sino al punto da potersi confondere con il risultato automatizzato.

Quello a cui si assiste è, quindi, un evidente passaggio di potere, ancorché parziale e, ad oggi, riferibile alla specifica fase processuale del *sentencing*. Non si tratta però, come spesso frettolosamente si adduce, un passaggio da una intelligenza umana (il giudice) a una artificiale (l'algoritmo), bensì da una umana (pubblica) e una altrettanto umana (quasi sempre privata), che trascina con sé, come si vedrà, notevoli e ulteriori preoccupazioni.

dell'algoritmo», si v. A. SIMONCINI, *L'algoritmo incostituzionale; intelligenza artificiale e il futuro delle libertà*, in *Rivista di BioDiritto*, 1, 2019, p. 81, e già da tempo è invalsi l'espressione di *algorithmic authority*, v. per esempio, C. LUSTIG ET. AL., *Algorithmic Authority: the Ethics, Politics, and Economics of Algorithms that Interpret, Decide, and Manage*, in *Chi Conference*, 2016, pp. 1057-1062. Con il medesimo significato si parla anche di *Philosophy anchoring*, si cfr. M. E. DONOHUE, *A replacement for justitia's scales?: machine learning's role in sentencing*, in *Harvard Journal of Law & Technology*, 32(2), pp. 657-678.

⁷⁶ La conclusione non muta neppure obiettando che in alcuni casi (si v., per esempio, *Samsa e J.S.*) il giudice non ha tenuto conto dell'indice di recidiva prodotto dalla macchina: e ciò sia poiché, come anche si vedrà in seguito, sono casi rari rispetto a quelli in cui il giudice assume la decisione sulla base dell'algoritmo; sia avendo in mente l'impianto culturale del modello giudiziario americano, in cui il *prosecutor* (il nostro pubblico ministero) possiede uno strapotere processuale, avendo nella propria disponibilità la polizia, le perizie medico legali, le prove, la possibilità di sentire i testimoni e dove le giurie sono spesso colpevoliste; in siffatto sistema, consegnare all'accusa un ulteriore elemento di prova (e di tale fascino, come l'algoritmo) rafforza ancora di più la disparità delle armi, ed è indicativo il fatto che i casi rinvenuti in cui il giudice ha trascurato il risultato algoritmico sono quelli in cui questi ignorava l'indice di recidiva basso o molto basso, applicando una pena di ammontare elevato (si cfr. ancora *Samsa e J.S.*). Sull'argomento si cfr. I. J. PATRONE, *Il prosecutor negli Stati Uniti. Un esempio da seguire?*, in *Questione Giustizia*, 11 marzo 2021.

7. Esternalizzazione della discrezionalità giudiziaria e opacità dell’algoritmo: una insidiosa “relazione biunivoca”

«[U]sing COMPAS to make a sentencing decision risks transferring sentencing discretion from judges to software developers»⁷⁷.

Una delle contestazioni sollevate dagli imputati in *People v. Younglove* concerneva proprio il rischio che la discrezionalità nella decisione da assumere nel *sentencing* si «trasferisse» dal giudice al progettista del *software* utilizzato nel processo.

Nonostante i giudici di appello non prendano seriamente tale doglianza, se non con affermazioni che ricalcano le distinzioni che abbiamo evidenziato nel precedente paragrafo, essa ci consente di introdurre un punto decisivo della riflessione fin qui svolta.

Gli algoritmi predittivi utilizzati nel sistema giudiziario statunitense sono, oggi, quasi esclusivamente progettati da società private⁷⁸ (COMPAS ne è un esempio) che, come tali, perseguono un interesse difficilmente coincidente con l’interesse della collettività.

Si assiste dunque a un primo fenomeno meritevole di considerazione: la progettazione di algoritmi decisionali che assumono un peso notevole, se non determinante, per la sorte degli imputati non è riservato all’amministrazione pubblica o, comunque, a organismi afferenti al potere giudiziario, ma viene, invece, *esternalizzata* a soggetti privati⁷⁹. Anche se ci si arrestasse qui, ci si potrebbe chiedere se già tale fenomeno possa ritenersi compatibile con i principi dello stato di diritto e della legittimazione democratica, seppure peculiare, della magistratura⁸⁰.

Ma, compiendo un passo ulteriore, il dubbio si rafforza e si assiste a un secondo fenomeno: abbiamo visto che l’impiego dell’algoritmo tende ad assorbire la stessa decisione discrezionale del giudice, il quale tende, anche non del tutto coscientemente, ad aderire a quanto l’algoritmo predittivo gli “suggerisce”. Stando così le cose, non sembrerebbe azzardato asserire che l’esternalizzazione della progettazione

⁷⁷ *People v. Younglove*, cit.

⁷⁸ Il fenomeno è mondiale, si cfr. T. SOURDIN, *Judges, Technology and Artificial Intelligence. The Artificial Judge*, cit., p. 273, secondo cui «[i]nternationally, much of the app and bot development in the justice sector is currently taking place in the private sector and is unregulated». Occorre comunque segnalare l’imponente riforma in corso negli Stati Uniti attuata, sul piano federale, con il *First Step Act* del 21 dicembre 2018, che, tra le altre cose, ha autorizzato il Procuratore Generale degli Stati Uniti a progettare un nuovo *tool*. Quest’ultimo, annunciato nel 2019 e nominato PATTERN, è uno strumento predittivo istituzionale e quindi non coperto da segreto ma, ciò nonostante, non mancano le critiche e gli interrogativi rispetto al suo funzionamento sollevati sia in dottrina sia in giurisprudenza. Si cfr., D. ZINGALES, *Risk assessment: una nuova sfida per la giustizia penale? La pericolosità criminale al vaglio algoritmico delle probabilità nell’esperienza statunitense*, cit., pp. 9-13 e B. L. GARRET - J. MONAHAN, *Judging Risk*, cit., pp. 484-486. In giurisprudenza si cfr. *Michael v. United States*, 2020.

⁷⁹ Si v. *mutatis mutandis*, con riguardo all’amministrazione per algoritmi, A. DI MARTINO, *L’amministrazione per algoritmi ed i pericoli del cambiamento in atto*, in *Il diritto dell’economia*, vol. 3, 2020, pp. 627 ss.

⁸⁰ Si è giustamente notato in dottrina che «*taking the human out of the equation faces risks of democratic legitimacy. This itself may lead to a reduction of judicial power. Additionally, with reduced democratic legitimacy, it is difficult to imagine that judicial discretion would remain a decisive factor in many cases*», si v. C. WINTER, *The challenges of artificial judicial decision-making for liberal democracy*, Working Paper n. 3 del 2021, p. 23.

dell'algoritmo si traduce, a ben vedere, in una parziale esternalizzazione, a un privato, della stessa funzione giudicante⁸¹, quantomeno nella specifica fase processuale del *sentencing*.

L'attribuzione formale del potere in capo al giudice penale diventa così un guscio vuoto che fatica a confinare una decisione assunta, di fatto, da chi ha progettato il *software*. In altre parole, la discrezionalità giudiziaria viene ineludibilmente attratta dal *software* e, quindi, dal tecnico progettista.

Il fenomeno della sostanziale esternalizzazione della funzione giurisdizionale si lega a doppio filo con un secondo elemento che la dottrina ha diffusamente analizzato e su cui i giudici americani, con evidente preoccupazione, hanno dovuto ragionare: l'opacità algoritmica (intesa, come vedremo, in termini diversi e tra loro complementari).

Gli algoritmi implementati nei *tools*, soprattutto quelli che fanno utilizzo di tecniche di *machine learning*, il più delle volte non sono trasparenti: anche senza voler utilizzare forzatamente il termine *black box*, ormai abusato nel dibattito giuridico e comunque non sempre correttamente impiegato⁸², per gli algoritmi più moderni (quelli cioè di intelligenza artificiale nella sua forma più evoluta, cioè, ad oggi, proprio il *machine learning*) non è sempre facile verificare il processo matematico che dagli *input* conduce a un *output* nonché risalire agli stessi dati di *training*, che spesso non vengono svelati⁸³.

Siffatte caratteristiche, a fronte della scarsa sensibilità dei giudici che appaiono ancora poco avvezzi ad apprezzarne le storture, meritano di essere brevemente descritte.

Si chiede il Giudice supremo dello *Iowa*, con notevole enfasi: «*How are we to determine the due process implications of the district court's use of risk assessment tools, when we do not know anything about the tools and Gordon failed to object to their use?*»⁸⁴.

Tali preoccupazioni inducono a indugiare sulla trasparenza non dal punto di vista dei diritti dell'imputato ad essa connessi – l'indagine dottrinale, come già indicato, prende infatti quasi sempre le mosse dalle problematiche che l'opacità algoritmica pone rispetto ai diritti di difesa, di accesso al giudice, di uguaglianza⁸⁵ – ma dalla prospettiva del giudicante che deve “fare i conti” con sistemi algoritmici di cui non conosce pressoché nulla e che fondano (parzialmente) il suo giudizio.

Se infatti, come si afferma in un passo della sentenza *Loomis*, «*[t]he circuit court and Loomis had access to the same copy of the risk assessment*» non è sufficiente porre uno sguardo critico sulle conseguenze dell'opacità

⁸¹ Si v. ancora, A. DI MARTINO, *L'amministrazione per algoritmi ed i pericoli del cambiamento in atto*, cit., p. 629.

⁸² Quantomeno in Italia, ormai si usa il termine *black box* per descrivere qualunque procedimento algoritmico, quando in realtà il termine andrebbe affibbiato ai soli algoritmi di intelligenza artificiale che utilizzano tecniche evolute di apprendimento automatico.

⁸³ Gli studi sulla opacità algoritmica non mancano. Si cfr., *ex multis*, C. RUDIN - C. WANG - B. COKER, *The Age of Secrecy and Unfairness in Recidivism Prediction*, in *Harvard Data Science Review*, Issue 2.1, 2020; J. VILLASENOR - V. FOGGO, *Artificial intelligence, due process, and criminal sentencing*, cit., spec. pp. 343 ss.

⁸⁴ *State v. Gordon*, 921 N.W.2d 19, 2018, sentenza che ribalta, come visto, la precedente pronuncia di appello. Si v. *supra* § precedente.

⁸⁵ Si cfr. la dottrina citata *supra* alla nota 2.

algoritmica con riguardo ai soli *diritti* del reo, ma, di riflesso, occorre soffermarsi anche su quelle che destano maggiore preoccupazione se scrutate dalla prospettiva del giudice.

Infatti, l'esternalizzazione del potere discrezionale, nell'accezione che qui si è sostenuta, non è che una conseguenza del mancato controllo che il giudice dovrebbe poter esercitare sull'algoritmo predittivo, in quanto sostanzialmente inaccessibile; detto altrimenti, l'esercizio del potere discrezionale non può venire correttamente esercitato se il giudice non ha piena contezza della "fonte" da cui la sua decisione promana, come accade, appunto, di fronte ad algoritmi opachi: da qui la "relazione biunivoca" tra esternalizzazione e opacità, dove quest'ultima da un lato è una concausa della esternalizzazione e a sua volta ne aggrava gli effetti.

Seppur scontando qualche inevitabile semplicismo, è possibile riconnettere l'opacità algoritmica a due distinti fattori: da un lato il fatto che la quasi totalità dei *tools* negli Stati Uniti, come visto, sono prodotti da società private che sono inevitabilmente restie e rivelare il codice algoritmico⁸⁶, normalmente protetto dal segreto industriale, dall'altro il funzionamento stesso dell'apprendimento automatico che, come detto, spesso non consente di verificare né i dati effettivamente utilizzati, né il peso assegnato a un singolo fattore.

Ora, se il primo problema può venire superato da obblighi di *disclosure* – siano essi legislativi o determinati da provvedimenti assunti dal giudice in sede processuale⁸⁷ – appare più complicato, oggi, fornire soluzioni soddisfacenti per fronteggiare la complessità procedimentale dell'IA e, in ultimo, per scongiurare l'applicazione di sentenze basate su informazioni inaccessibili⁸⁸.

Tali preoccupazioni trovano ancoraggio, ancora una volta, nelle parole dei giudici statunitensi che sono stati chiamati a fare utilizzo dei sistemi algoritmici di predizione del rischio.

⁸⁶ Ma si v. *supra*, nota 78.

⁸⁷ Si segnala che, di recente, una Corte distrettuale di New York ha obbligato Northpoint (ora Equivant), la società produttrice di COMPAS, a svelare i dati contenuti nell'algoritmo, bilanciando tale obbligo attraverso un *Protective Order* che impedisce a chi esamina i dati la diffusione o riproduzione del materiale. Si cfr. nella pronuncia *Flores v. Stanford*, WL 4441614, 2021.

⁸⁸ In questo senso, di interesse è anche il caso *State v. Buford* (2019), che offre un esempio interessante di come l'atteggiamento dei giudici americani di fronte a strumenti di analisi algoritmica non trova riscontro in pronunce che, non occupandosi direttamente dell'utilizzo di COMPAS o di altri strumenti di analisi algoritmica ma di situazioni senz'altro analoghe, sembrano contraddire apertamente la consolidata giurisprudenza americana sui *risk assessment tools*. Nella pronuncia in esame la Corte d'appello del Wisconsin annulla la sentenza di primo grado e rinvia per una nuova determinazione della pena da applicare in concreto in quanto la Corte distrettuale aveva basato la propria decisione su un'«informazione inaccurata» e in particolare su uno studio statistico sull'incidenza degli abusi sessuali sui minori, dal quale emergeva un indice particolarmente alto. In questo caso la Corte d'appello, condividendo le doglianze del condannato, riteneva che quel dato non fosse presente nello studio utilizzato dal giudice di primo grado e quindi non poteva costituire un elemento su cui fondare l'applicazione della pena. Alla controdeduzione dello Stato secondo cui, anche se la Corte si era basata su quella informazione erronea, la decisione sarebbe stata la medesima, la Corte replica che «*it is not clear beyond a reasonable doubt that the circuit court would have imposed the same sentence without this inaccurate information*».

Nella decisione *State v. Guise*⁸⁹, coeva alla *Gordon* (si cfr *supra* § 6) la Corte d'appello dello Iowa riscontra un abuso di discrezionalità da parte del giudice distrettuale, il quale aveva assunto la propria determinazione tenendo conto delle risultanze del *risk assessment tool* (denominato IRR). A sostegno della propria decisione, in maniera indiscutibilmente decisa, i giudici di appello affermano: «*[w]e do not know what the IRR is, what factors led to the recommendation of intensive supervision, or whether the factors were appropriate for consideration in the sentencing context. It is impossible to determine whether the IRR was relevant to the question of sentencing within the meaning of section 901.2(1)* [si tratta di una sezione dello *Iowa Code*]».

Inoltre, nel tentativo di valutare a fondo le caratteristiche del sistema e la sua finalità, la Corte precisa che non vi era traccia di come IRR catalogava le informazioni, come assegnava i punteggi, quali fattori venivano considerati e come venivano pesati: «*[w]e simply do not know*», affermano perentoriamente i giudici⁹⁰.

In altre parole: «*[t]he IRR as described in Guise's PSI report was a black box, devoid of transparency*» e, mancando le informazioni essenziali che consentono alla Corte (e non solo all'imputato!) di valutare l'affidabilità dell'IRR, «*the court's use of the test in the sentencing decision amounted to an abuse of discretion*»⁹¹.

L'opacità come fattore che indirizza e, allo stesso tempo, ostacola il corretto esercizio del potere discrezionale diventa argomento di discussione di recente anche nel Michigan⁹². In *People v. Canedo* 2021⁹³, la *concurring opinion* della Giudice Cavanagh alla decisione della *Supreme Court* del Michigan coglie un punto chiave: se è vero, afferma la giudice, che l'imputato ha diritto a essere giudicato sulla base di «*accurate information*»⁹⁴, «*in the context of COMPAS risk assessments, it is unclear to me what it might mean to measure the accuracy of a prediction about an individual's future conduct and how that prediction might be challenged without knowing how it was formulated*».

Di fronte a un algoritmo come COMPAS che è «*proprietary, and undisclosed*»⁹⁵, è dunque compito arduo per un giudice esercitare correttamente il proprio potere discrezionale e adempiere *tout court* alla sua funzione:

⁸⁹ *State v. Guise*, 919 N.W.2d 635, 2018. Il caso riguardava la condanna di Guise per il reato di furto con scasso a una pena superiore a dieci anni di carcere, assunta anche sulla base dello *score* prodotto dal sistema algoritmico IRR.

⁹⁰ Già in dottrina si segnalava l'esigenza che il giudice non potesse arrendersi all'opacità ma che dovesse richiedere, per ritenere compatibile l'utilizzo dell'algoritmo con le garanzie istituzionali, una *giustificazione* del risultato. In questo senso si cfr A. DEEKS, *The judicial demand for explainable artificial intelligence*, in *Columbia Law Review*, vol. 119, n.7, p. 1843 e pp. 1845 ss.

⁹¹ La sentenza, come accaduto in *Gordon*, viene poi annullata dalla *Supreme Court* dello Iowa, si cfr. *State v. Guise*, 921 N.W.2d 26, 2018. Dopo le pronunce della Corte suprema dello Iowa (in *Guise* e in *Gordon*), in tale Stato si segnalano, di recente, diverse pronunce che hanno sempre respinto gli appelli di soggetti che lamentavano la violazione del giusto processo in virtù della *reliance* del giudice su sistemi algoritmici di valutazione del rischio: si cfr. *State v. Houghmaster*, 946 N.W.2d 770, 2020; *State v. Jacobson*, 940 N.W.2d 448, 2019.; *State v. O'Connor*, 942 N.W.2d 4, 2020; *State v. Mollo*, 928 N.W.2d 899, 2019.

⁹² Nel cui Stato abbiamo già incontrato la sentenza *People v. Younglove*, presa in esame *supra* al § 5.

⁹³ *People v. Canedo*, 507 Mich. 1029, 2021.

⁹⁴ Cfr. ancora *Buford*, menzionato *supra* alla nota 88.

⁹⁵ Si v., ancora, la *concurring opinion* della Giudice Cavanagh in *People v. Canedo*.

da un lato verrebbe meno qualunque affidabilità di una sentenza fondata su dati “invisibili” (al riguardo, è indicativo che si discuta diffusamente di come migliorare la *accuracy* dei sistemi algoritmici, ma nulla non si dice sulla oggettiva impossibilità di considerare *accurate*, e quindi corretta, una decisione fondata su dati che, per quanto affidabili, non sono svelati in sede processuale), dall’altro, come già anticipato, ancor più sarebbe pregiudicata la legittimazione stessa del potere giudiziario, la cui decisione viene attratta e assorbita da un algoritmo non solo privato, ma anche impenetrabile.

In sintesi: l’impiego di sistemi algoritmici comporta la tendenza del giudice ad affidare ad essi il proprio giudizio, la cui rilevanza nella decisione sulla determinazione della pena diviene preponderante; tale assunto, legato al fatto che gli algoritmi impiegati non nascono nel circuito pubblico ma sono progettati da soggetti privati, comporta che una parte rilevante del potere discrezionale del giudice venga esternalizzato ai soggetti privati che progettano i *software*; il fenomeno dell’esternalizzazione è aggravato (se non proprio causato, come visto) dall’impossibilità dei giudici – sia per loro carenze sia per le caratteristiche degli algoritmi di *machine learning* – di conoscere a fondo quali strumenti stanno utilizzando, costringendoli pertanto a rinunciare al loro ruolo di custodi ultimi dei diritti, tanto più in un ambito – come è quello penale – dove l’assunzione di una decisione quanto più possibile oculata, trasparente, “giusta”, assume un rilievo decisivo.

Giunti sino a qui è lecito chiedersi se l’abuso di discrezionalità, che solo in rari casi è stato riscontrato dai giudici americani ma che assume, come visto, contorni più estesi, potrebbe in effetti alleviarsi tramite un utilizzo accurato dello strumento principe di verifica critica, da parte tanto dell’imputato quanto della pubblica opinione, dell’esercizio della discrezionalità, vale a dire la motivazione. In questi termini, come si dirà, traluce un’insufficienza intrinseca dello sforzo motivazionale, sia per le caratteristiche proprie dei moderni sistemi algoritmici di *risk assessment* – come già ampiamente evidenziati in questo paragrafo – sia per il divario, oggi ancora ampio, tra la complessità di siffatti sistemi e la formazione culturale del giudice (v. *infra*, § 10).

8. La motivazione come argine all’*abuse of discretion*?

Si legge in *Loomis*: «[a]n erroneous exercise of discretion occurs when a circuit court imposes a sentence “without the underpinnings of an explained judicial reasoning process”»⁹⁶

Quando mancano i “sostegni” in un processo di ragionamento motivato, quindi, non può che ravvisarsi un esercizio fallace del potere discrezionale, privato degli elementi essenziali per il suo costruirsi. La soluzione cui giungono in modo non del tutto convincente i giudici del Wisconsin, tuttavia, recupera la legittimità dell’esercizio motivazionale nella non esclusività del *tool* utilizzato, specificando, tuttavia, che

⁹⁶ *Loomis*, § 30.

«a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed»⁹⁷.

Che la motivazione costituisca, in particolar modo nel processo penale, una garanzia ordinamentale irrinunciabile tanto endoprocessuale quanto extraprocessuale è cosa nota⁹⁸. Occorre, pertanto, verificare se per contenere il fenomeno della esternalizzazione sia sufficiente, come indicato dal giudice americano, imporre al giudice uno sforzo motivazionale che valorizzi fattori diversi e indipendenti rispetto al punteggio fornito dall’algoritmo, o se anche questa conclusione sia in realtà solo un ulteriore velo di Maya che nasconde una realtà più complessa.

A tal proposito preme rammentare che il discorso sino a qui sviluppato muove dalle considerazioni svolte dai giudici americani i quali, non di rado, pur negando che il potere discrezionale da parte dei colleghi di prime cure fosse esercitato scorrettamente, hanno comunque manifestato, come visto, un notevole disagio di fronte all’inaccessibilità (*recte* incomprendibilità) dell’algoritmo (si v. *supra*, § 7, spec. *Gordon e Guise*).

Siffatta premessa si rende necessaria in quanto un – seppur sintetico – affondo sulla motivazione non può che prendere le mosse da un fondamentale *distinguo*: altro è delineare il perimetro dell’obbligo motivazionale a fronte di algoritmi del tutto trasparenti o, perlomeno, facilmente comprensibili dalle parti e dal giudice, altro è considerare legittima una motivazione che si fonda, parzialmente o esclusivamente, su un’informazione prodotta dagli algoritmi di quarta (o quinta, a seconda della classificazione prescelta) generazione, che rimangono tuttora oscuri agli occhi del giudice.

Nella prima ipotesi nulla osta a che, attraverso un impiego accurato della motivazione – che tuttavia potrà raggiungersi solo allorché il giudice sarà pienamente consapevole dello strumento utilizzato (si v. *infra*, § 10) – la discrezionalità venga ad esercitarsi in modo legittimo: è facilmente immaginabile un giudice che, ben conscio dei dati inseriti all’interno dell’algoritmo e delle modalità attraverso cui esso produce il punteggio di rischio, consideri quest’ultimo alla stregua degli altri elementi di fatto emersi in sede processuale, confermando ovvero disattendendo il risultato algoritmico attraverso una motivazione che, secondo autorevole dottrina, dovrebbe essere *rafforzata*⁹⁹. In tal senso, qualora un algoritmo smentisca una valutazione di pericolosità autonomamente effettuata dal giudice, quest’ultimo avrebbe l’obbligo di «confutare con argomentazioni di “efficacia scardinante” o di “forza persuasiva superiore” questa *second opinion* “artificiale”»¹⁰⁰.

⁹⁷ *Loomis*, § 99.

⁹⁸ S. LORUSSO, *Il diritto alla motivazione*, in *Diritto penale contemporaneo*, 2018.

⁹⁹ Si cfr. V. MANES, *L’oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, cit., pp. 19-22.

¹⁰⁰ Si cfr. V. MANES, *op cit.*, p. 20. Si veda anche M. CATERINI, *Il giudice penale robot*, in *La legislazione penale*, 2020, p. 17, in cui l’A. scrive «A questa *pars costruens* della motivazione dovrebbe affiancarsi una *destruens*, vale a dire la confutazione, ossia l’esplicitazione delle ragioni per le quali l’opzione interpretativa più favorevole è implausibile o illogica e, come tale,

Del resto, proprio nell'ordinamento italiano si ritrova la riproposizione di un siffatto approccio nella giurisprudenza amministrativa che ha meritoriamente ricostruito il rapporto tra decisore umano e algoritmo, ancorché nell'ambito della decisione amministrativa e non della decisione giudiziale¹⁰¹. Ciononostante, i principi posti in luce dai Tribunali amministrativi regionali prima, e dal Consiglio di Stato poi, ben riflettono ciò che qui si intende segnalare.

Infatti, secondo la più recente giurisprudenza amministrativa la legittimità dell'impiego dell'algoritmo in sede decisionale è subordinata alla *conoscibilità* dello stesso, a ciò non essendo sufficiente il mero accesso al codice sorgente dell'algoritmo – come aveva sostenuto il T.A.R. nel 2017¹⁰² e poi ancora nel 2018¹⁰³ – essendo necessaria la predisposizione di spiegazioni che traducano «la formula algoritmica in una regola giuridica che sia «leggibile e comprensibile, sia per i cittadini che per il giudice»¹⁰⁴. Detta regola algoritmica, peraltro, deve rimanere soggetta alla piena cognizione del giudice amministrativo, quale inevitabile «declinazione diretta del diritto di difesa del cittadino, al quale non può essere precluso di conoscere le modalità (anche se automatizzate) con le quali è stata in concreto assunta una decisione destinata a ripercuotersi sulla sua sfera giuridica»¹⁰⁵.

In continuità con quanto affermato, i Supremi Giudici amministrativi, in alcune sentenze successive¹⁰⁶, mettono ulteriormente a punto lo 'statuto costituzionale' dell'algoritmo: innanzitutto la conoscibilità è da intendersi come piena *comprensibilità* dell'algoritmo, non potendosi dare rilievo alla «riservatezza delle imprese produttrici dei meccanismi informatici utilizzati i quali, ponendo al servizio del potere autoritativo tali strumenti, all'evidenza ne accettano le relative conseguenze in termini di necessaria trasparenza»¹⁰⁷. Inoltre, in armonia con quanto affermato dai giudici americani, si afferma il principio di

da escludere dal ventaglio delle alternative ermeneutiche a disposizione per la decisione. dovrebbe spiegare *funditus* anche le ragioni per le quali risulti implausibile l'opzione interpretativa più favorevole suggerita dal sistema esperto».

¹⁰¹ I commenti dottrinali alla giurisprudenza amministrativa che qui si menziona si sprecano. Tra i tanti, si vedano, per la lucidità delle argomentazioni, D.U. GALETTA, *Algoritmi, procedimento amministrativo e garanzie: brevi riflessioni, anche alla luce degli ultimi arresti giurisprudenziali in materia*, in *Rivista Italiana di Diritto Pubblico Comunitario*, 3, 2020, pp. 501 ss.; P. OTRANTO, *Riflessioni in tema di decisione amministrativa, intelligenza artificiale e legalità*, in *Federalismi.it*, 10 marzo 2021. B. MARCHETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *BioLaw Journal – Rivista di BioDiritto*, n. 2/2021; L. MUSSELLI, *La decisione amministrativa nell'età degli algoritmi: primi spunti*, in *MediaLaws-Rivista di diritto dei media*, 1/2020; S. VERNILE, *Verso la decisione amministrativa algoritmica*, in *MediaLaws-Rivista di diritto dei media*, 2/2020.

¹⁰² T.A.R. Lazio, sent., Sez. III-bis, n. 3742 e 3769 del 2017. In commento alle sentenze v. E. PROSPERETTI, *Accesso al software e al relativo algoritmo nei procedimenti amministrativi e giudiziali. un'analisi a partire da due pronunce del TAR Lazio*, in *Dir. informaz. e inf.*, 4-5, 2019, pp. 979 e ss.

¹⁰³ T.A.R. Lazio, Sez. III-bis, sentt. nn. 9924-9230 del 2018 su cui si v. *ex multis*, A. SIMONCINI, *Profili costituzionali della amministrazione algoritmica*, in *Riv. Trim. Dir. Pubbl.*, 4, 2019, p. 1162.

¹⁰⁴ Cons. Stato, Sez. VI, sent. n. 2270 del 2019.

¹⁰⁵ *Ibidem*.

¹⁰⁶ Cons. Stato, Sez. VI, sent. n. 8472 del 2019 e sent. n. 881 del 2020.

¹⁰⁷ *Ibidem*.

non esclusività della decisione algoritmica, ricavato dall'art. 22 dal *Gdpr*, cui si ricollega la necessaria imputabilità della decisione algoritmica al funzionario pubblico¹⁰⁸.

Solo alla trasparenza così intesa e alla conseguente permanenza dell'imputabilità della decisione all'essere umano può coesistere un reale obbligo di motivazione che, come visto, si rafforza, avendo l'amministrazione l'obbligo di rendere comprensibile la propria decisione¹⁰⁹.

Come si vede, quindi, secondo la giurisprudenza amministrativa il tassello irrinunciabile per verificare la legittimità ovvero l'illegittimità di un provvedimento amministrativo assunto (anche) da un algoritmo è la *comprensibilità* dello stesso, secondo una declinazione *rafforzata* del principio di trasparenza, cui fa da corollario l'obbligo di motivazione, che rende la decisione compatibile con l'imputazione diretta della decisione al funzionario pubblico nonché con le garanzie processuali del destinatario del provvedimento (in particolare il diritto di difesa e il diritto di accesso al giudice).

Se dunque la premessa per garantire che la discrezionalità dell'amministrazione venga correttamente esercitata, attraverso un attento e rafforzato obbligo motivazionale, è la piena padronanza dell'algoritmo impiegato, viene da chiedersi se l'approccio seguito dalla giurisprudenza italiana sia trapiantabile *sic et simpliciter* per giustificare l'esercizio discrezionale del potere da parte del giudice penale di fronte ai *risk assessment tools*, soprattutto quelli più sofisticati ed evoluti.

La risposta parrebbe essere negativa.

Come visto nei casi americani esaminati non è vero che manca una motivazione *ai fini* della sindacabilità della decisione, tant'è che affermazioni dei giudici sono tratte da sentenze di secondo grado chiamate a sindacare l'esercizio abusivo della discrezionalità da parte dei giudici di prime cure che la difesa degli imputati aveva tentato di censurare.

E tuttavia non può che riconoscersi l'insufficienza dello sforzo motivazionale dei giudici per due ordini di ragioni: in primo luogo l'inaccessibilità degli algoritmi più evoluti implementati negli strumenti dei *risk assessment* non consente al giudice di rendere comprensibile – e quindi legittima – la propria decisione, e ciò sia nel caso in cui decida di disattendere il risultato automatico, assegnando peso decisivo ad altri fattori indipendenti, sia che invece assuma il punteggio di rischio a fondamento della commisurazione di pena; in secondo luogo, anche nei casi di impiego di algoritmi *astrattamente* idonei a consentire l'accesso al codice e di rendere noti i dati e i processi di determinazione del punteggio di rischio, sembra mancare

¹⁰⁸ Anche se non rileva direttamente ai fini del presente lavoro, i giudici amministrativi enunciano altresì il principio di non discriminazione algoritmica, secondo cui, anche di fronte ad algoritmi perfettamente conoscibili e non costituenti l'esclusiva fonte della decisione finale, essi non devono produrre effetti discriminatori nei confronti dei destinatari, si cfr. ancora le sentenze da ultimo citate.

¹⁰⁹ In una recentissima decisione del T.A.R. Campania si legge che «il fatto che il provvedimento venga emanato sulla scorta di una complessa operazione di calcolo produce l'opposto effetto di rafforzare, per certi versi, l'obbligo motivazionale in capo all'Amministrazione, la quale dovrà rendere la propria decisione finale non solo conoscibile, ma anche comprensibile», cfr. T.A.R. Campania, Sez. III, sent. n. 7003 del 14 novembre 2022.

al giudice, come meglio si chiarirà *infra* (§ 10) la formazione professionale adeguata per governare al meglio una decisione che rischia di risultare intrinsecamente *immotivata* e quindi illegittima.

Sembrerebbe quindi che l'assolvimento di un vero e legittimo onere motivazionale, che permetta di addivenire a un esercizio discrezionale che sia coerente con la funzione del giudice, potrà raggiungersi solo allorché gli algoritmi saranno realmente accessibili e che gli stessi giudici acquisiscano gli strumenti adeguati al fine di ponderare correttamente tutti gli elementi di fatto rilevanti nel processo. Tale impostazione sembra del resto coerente con i primi tentativi, in Europa, di regolare l'utilizzo di algoritmi predittivi all'interno dei sistemi giudiziari, il cui breve cenno è essenziale al fine di svolgere qualche considerazione più generale che travalichi i confini americani e orienti il dibattito anche all'interno nostro ordinamento.

9. Uno sguardo sull'Europa: i primi embrionali tentativi di regolare l'intelligenza artificiale all'interno dei sistemi giudiziari

Dall'analisi dei casi americani si è raggiunta una prima conclusione: se l'obiettivo dell'introduzione dei *risk assessment tools* nel procedimento penale perseguiva l'obiettivo di *razionalizzare* la discrezionalità del giudice, si è visto che tale (incompiuta) razionalizzazione si è accompagnata a un disseccamento del potere discrezionale del giudice e al suo parziale spostamento dell'autorità pubblica al soggetto privato¹¹⁰. Si è poi tentato di dimostrare che siffatto fenomeno, quanto meno nell'ipotesi di *tools* evoluti, difficilmente può venire contrastato attraverso un uso accorto della motivazione.

A questo punto deve risultare chiaro che il tentativo di oggettivazione della discrezionalità non rischia di indebolire, come si è autorevolmente affermato, l'«aspetto umano della decisione»¹¹¹, poiché l'essere umano non svanisce (rimanendo presente nella persona del progettista); il rischio che si corre è semmai quello di svigorire la stessa idea di giurisdizione intesa come la competenza dell'autorità pubblica, nella specie il giudice, di “dire il diritto” (*ius dicere*) e di “giustificare” (*iustum facere*) nel doppio valore semantico di motivare adeguatamente la decisione e, in ultimo, di fare giustizia.

Quali dunque le possibili soluzioni per arginare il fenomeno dell'esternalizzazione e riconsegnare al giudice penale le chiavi della decisione¹¹²?

¹¹⁰ Sulle difficoltà di razionalizzare la discrezionalità tramite IA si cfr. uno scritto precursore di M. TARUFFO, *Judicial Decisions and Artificial Intelligence*, in *Artificial Intelligence and Law*, 1998, pp. 311-324, spec. p. 321.

¹¹¹ A. BARAK, *La discrezionalità del giudice*, Giuffrè, Milano, 1995, p. 130.

¹¹² Il contributo ha preso le mosse da una specifica fase processuale in cui sembra che la discrezionalità del giudice si svuoti. In realtà, il discorso può ben venire esteso anche per situazioni diverse e in campi diversi, situazioni in cui il giudice, per giungere alla propria decisione, deve considerare un risultato prodotto da un algoritmo predittivo che in qualche modo orienta la medesima decisione.

Sul versante dell'algoritmo, fino a che le tecniche più evolute di *machine learning* rimarranno poco comprensibili e incapaci di spiegare il procedimento che da determinati dati perviene a un *output*¹¹³, non sembra possibile ammetterne l'utilizzo in un processo penale¹¹⁴. Le promesse di celerità, affidabilità e certezza degli algoritmi impiegati nel processo, soprattutto nel processo penale, devono recedere di fronte alla necessità che la giustizia si mantenga all'interno di binari di trasparenza e pubblicità.

Non dissimile appare il giudizio su algoritmi (cd. deterministici¹¹⁵) meno sofisticati, teoricamente (e solo teoricamente) compatibili con le esigenze di comprensibilità, una volta svelati i codici e i dati attraverso cui essi operano. E ciò in quanto, come si è visto, l'opacità algoritmica deriva da due distinti fattori, uno propriamente tecnico – la *black box* che il *machine learning* inevitabilmente genera – e uno legato all'“industria” dell'intelligenza artificiale, vale a dire il monopolio privato nella progettazione di algoritmi che poi vengono usati per l'amministrazione della giustizia. Tale fenomeno comporta quindi – anche per algoritmi le cui caratteristiche tecniche ne permetterebbero la comprensione – la frequente inaccessibilità degli stessi, che rimangono protetti da norme sulla proprietà industriale e dall'indisponibilità delle società produttrici di svelare dati e codici sorgenti degli algoritmi.

Inoltre, per la ragioni che abbiamo tentato di delineare nel presente lavoro, il rischio è quello di esternalizzare gran parte della grammatica processuale (compresa la decisione finale) e, in generale, la

¹¹³ Si segnala che la ricerca scientifica oggi sta concentrando l'attenzione su due strade distinte: l'*explainability* dei sistemi di IA (anche detta XAI, *explainable artificial intelligence*) che permette di comprendere le variabili e i procedimenti interni degli algoritmi, e l'*interpretability* – meno ambiziosa – che consiste di valutare il sistema sulla base dell'*output* finale, senza necessariamente comprendere ogni singolo passaggio del procedimento algoritmico. Una strada che di recente sta venendo intrapresa è l'uso di scenari controfattuali: il modello selezione campioni alternativi con caratteristiche diverse, confronta i diversi risultati e identifica come e perché differiscono. Per esempio, il modello sarà in grado di rilevare che il risultato sarebbe stato diverso se la caratteristica X fosse mancata o se quella Y fosse stata aggiunta, si v., *ex plurimis*, M. HARRADON - J. DRUCE - B. RUTTEMBERG, *Casual Learning and Explanation of Deep Neural Networks via Autoencoded Activations*, 2018. V. anche A. DEEKS, *The judicial demand for explainable artificial intelligence*, cit., pp. 1829 ss. Ciò detto, non è sempre possibile spiegare *ex post* i procedimenti algoritmici, sia per il fatto che non sempre è facile identificare, in un modello *black box*, i fattori decisivi che hanno portato a un determinato risultato, sia perché anche qualora sia identificato il fattore causativo, non sempre può servire quale vera e propria spiegazione del procedimento. Si cfr., sul punto, M. SCHERER, *Artificial Intelligence and legal Decision-Making: the wide open?*, in *Journal of International Arbitration*, 36, 5, 2019, pp. 563 ss.

¹¹⁴ Peraltro nel nostro ordinamento vige ancora il divieto di perizia criminologica di cui all'art. 220, comma 2, c.p.p., che sembra costituire un importante ostacolo per l'inserimento di sistemi di predizione del rischio e la cui *ratio* è stata ricostruita dalla Corte costituzionale nella sentenza n. 124 del 1970 nella quale, sindacando la legittimità dell'art. 314 del codice di procedura penale del 1930 (riprodotto dall'art. 220 del codice vigente), la Corte ha giustificato la disposizione anche dando valore al fatto che «lo studio della personalità dell'imputato possa venir compiuto solo da chi abbia presente anche il carattere afflittivo e intimidatorio della pena, con cui la finalità di rieducazione deve essere temperata».

¹¹⁵ Nella letteratura giuridica vengono definiti anche “condizionali”, si cfr., per esempio, G. CARULLO, *Decisione amministrativa e intelligenza artificiale*, cit. p. 434. Sono algoritmi che utilizzano la cd. *Symbolic AI* (o *Human readable AI*), vale a dire algoritmi che si basano su regole stabilite dal programmatore e assumono decisioni sulla base di modelli predefiniti. A una serie di condizioni l'algoritmo produce il proprio *output* secondo una logica *if-then*. Il procedimento logico-inferenziale di tali algoritmi è dunque facilmente verificabile, essendo sempre possibile risalire alle condizioni (*id est* i dati) sui cui l'algoritmo ha prodotto la propria determinazione. Per un approfondimento si cfr. E. PERES, *Che cosa sono gli algoritmi*, Salani, Firenze, 2020.

potestà pubblica del giudice, a soggetti privati, da cui l'inaridirsi della discrezionalità che compete al soggetto pubblico decidente.

Da queste premesse pare doversi quindi orientare il percorso futuro verso la realizzazione pubblica, o comunque parzialmente rientrante nel circuito pubblico, di algoritmi predittivi e decisionali.

Per l'economia del presente lavoro non è possibile qui passare in rassegna ogni possibile soluzione che possa soddisfare tale esigenza: certamente non è peregrino pensare all'istituzioni di *pool* di funzionari pubblici, che operino a stretto contatto con i giudici, altamente preparati e in grado di progettare, in autonomia o con l'ausilio di soggetti e investimenti privati, algoritmi trasparenti e diretti esclusivamente a un impiego giurisdizionale.

Tali algoritmi potrebbero inoltre ricevere, da parte di organismi indipendenti, un'omologazione o certificazione¹¹⁶ che ne attesti l'affidabilità e la coerenza dei dati inseriti, e che ne garantisca l'aggiornamento costante.

In questo senso, del resto, sembrano muoversi anche le istituzioni europee, che ormai da qualche anno stanno monitorando attentamente le problematiche afferenti all'impiego di strumenti di intelligenza artificiale all'interno dei sistemi giudiziari, tentando di riallineare il fenomeno ai principi fondamentali che dovrebbero regolare la giurisdizione nel senso più ampio del termine.

Un ruolo da protagonista lo ha assunto il Consiglio d'Europa, che tramite la *European Commission for the efficiency of justice* (CEPEJ) ha avuto il merito di adottare il primo documento europeo teso ad affrontare organicamente il tema dell'IA nel processo. Con la *Ethical Charter on the use of artificial intelligence in judicial systems and their environment*¹¹⁷ la CEPEJ enuncia infatti per la prima volta i principi fondamentali che dovrebbero guidare l'utilizzo dell'IA nei sistemi giudiziari e fornisce un primo quadro dello stato dell'arte nei paesi del Consiglio d'Europa; tra i cinque principi enunciati nell'incipit del documento, il principio del *rispetto dei diritti fondamentali*, il principio di *non discriminazione*, il principio di *qualità e sicurezza*, il principio di *trasparenza, imparzialità ed equità* e il principio del *controllo da parte dell'utilizzatore*, gli ultimi tre assumono particolare rilevanza all'interno della riflessione qui svolta.

Il principio di qualità e sicurezza richiede l'instaurazione di un ambiente tecnologico sicuro, che deve essere perseguito attraverso un approccio quanto più possibile interdisciplinare: a tal fine, i progettisti di

¹¹⁶ Si cfr., di recente, U. RUFFOLO, *Giustizia predittiva, decisione algoritmica e ruolo ancillare della macchina come ausiliario necessario del giudice umano*, in A. PAJNO - F. DONATI - A. PERRUCCI, *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, Il Mulino, 2022, pp. 453-481.

¹¹⁷ I commenti alla Carta Etica sono sterminati. Si v. per tutti C. BARBARO, *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, in *Questione giustizia*, trimestrale n. 4, 2018, pp. 189 ss.; S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, fasc. 4, 2019, pp. 1748 ss.; 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; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti e Europa*, cit., pp. 12 ss.

algoritmi di *machine learning* dovrebbero fare affidamento sulle competenze di professionisti, sia in ambito giuridico (giudici, avvocati, pubblici ministeri...), sia nel campo delle scienze sociali quali economisti, sociologi e filosofi.

Il principio di trasparenza, imparzialità ed equità si sostanzia invece nella ricerca di un equilibrio tra i diritti di proprietà intellettuale sugli algoritmi – si pensi a COMPAS – e l’ineludibile necessità che il processo decisionale sia trasparente, imparziale ed equo. Si legge nella Carta etica che si dovrebbe assicurare la totale trasparenza tecnica degli algoritmi ovvero, qualora ciò non fosse possibile, sarebbe necessario quanto meno garantire la spiegazione, in un linguaggio comprensibile, del funzionamento dell’algoritmo e del procedimento seguito.

Il principio del controllo da parte dell’utilizzatore, infine, impone che il giudice debba essere in grado in ogni momento di rivedere la decisione assunta dalla macchina e di non essere vincolato a essa qualora il caso concreto richieda una decisione diversa.

Dai principi espressi nella *Ethical Charter* emerge un punto fondamentale, a conferma delle riflessioni qui svolte: il principio *under user control* viene indicato per ultimo nel documento, quasi a chiusura del sistema. Solo se gli altri principi vengono implementati, allora si riscopre il senso di concepire un giudice in grado di governare l’algoritmo e di assumersi la piena responsabilità della decisione: questo obiettivo è tuttavia raggiungibile, come vedremo nel successivo paragrafo, solo attraverso un rinnovamento professionale e culturale dell’apparato giudiziario.

Anche l’Unione europea si sta muovendo verso una regolazione organica (e vincolante) dei sistemi di IA impiegati nei sistemi giudiziari: nella proposta di regolamento sull’intelligenza artificiale (cd. *Artificial intelligence act*¹¹⁸), destinata – sembra – a costituire il punto di riferimento in materia nei prossimi anni, viene adottato il cd. *Risk based approach*, un approccio basato sul rischio che mira a fornire una forma di protezione preventiva, imponendo tanto più stringenti condizioni e requisiti quanto più rischiosi appaiono gli strumenti di IA prodotti. In questo senso, tra i sistemi di IA “ad altro rischio” la proposta ricomprende «i sistemi di IA destinati ad assistere un’ autorità giudiziaria nella ricerca e nell’interpretazione dei fatti e del diritto e nell’applicazione della legge a una serie concreta di fatti» (All. III, punto 8 (a))¹¹⁹, a conferma che gli strumenti utilizzati al di là dell’oceano sono ad oggi visti con grande preoccupazione.

¹¹⁸ Per alcuni commenti alla proposta si cfr., tra i tanti, L. FLORIDI, *The European Legislation on AI: A Brief Analysis of its Philosophical Approach*, in *Philos. Technol.*, 34, 2021, pp. 215-222; F. PIZZETTI, *La proposta di Regolamento sull’IA della Commissione Europea presentata il 21.4.2021 (COM (2021) 206 final) tra Mercato Unico e competizione digitale globale*, in *Diritto di internet*, n. 4/2021, pp. 591 ss.; C. CASONATO - B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell’Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, n. 3/2021, pp. 415-437.

¹¹⁹ Per siffatti sistemi il Regolamento impone l’apposizione della Marcatura CE per indicare la loro conformità alla normativa europea nonché per poter circolare liberamente nel mercato interno (si cfr. considerando n. 67 e art. 49 della Proposta).

L'AI Act è già stato oggetto di alcuni pareri istituzionali: particolarmente significativa l'*opinion* del Comitato economico sociale europeo (CESE) che, in riferimento proprio all'utilizzo della intelligenza artificiale nell'amministrazione della giustizia, afferma che esso rappresenta «*un tema particolarmente delicato e dovrebbe essere affrontato con più sfumature e controlli di quanto non avvenga attualmente. Il riferimento ai sistemi di LA utilizzati per assistere l'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione delle norme a una serie concreta di fatti non considera che giudicare è attività assai più complessa della mera ricerca di schemi ricorrenti in dati storici (che in sostanza è ciò che gli attuali sistemi di intelligenza artificiale fanno)*»¹²⁰. Il CESE riconosce pertanto l'importanza di non sottovalutare il fenomeno e ammonisce rispetto alla complessità dell'attività giurisdizionale che non può essere relegata al mero appiattimento su dati statistici e predittivi. I primi timidi passi compiuti dai legislatori europei non appaiono tuttavia da soli sufficienti a rischiarare il cielo nebuloso che copre la relazione tra giudici e algoritmi predittivi; la visione antropocentrica del diritto – che, per quanto talvolta messa in dubbio dall'evoluzione tecnologica, non pare possa arretrare in tempi brevi – impone di ricercare delle soluzioni non solo dal versante dell'algoritmo, ma allargando lo sguardo all'essere umano, nella specie al giudice.

Se anche si trovasse il modo di garantire la spiegabilità degli algoritmi di *machine learning* più evoluti; se anche introiettassero i problemi dell'esternalizzazione della funzione giurisdizionale e, quindi, si ripensasse alla progettazione dell'algoritmo da una prospettiva pubblica; se anche, in questo modo, si riuscissero a progettare algoritmi in grado di conformarsi alle garanzie costituzionali e processuali; se tutto ciò accadesse – e non è affatto detto che accada – rimarrebbe un profondo, e assorbente ostacolo, la attuale carenza di formazione “tecnologica” del giudice.

10. Alcuni spunti conclusivi: la formazione dei giudici e la ricerca di un nuovo principio processuale: *algorithmia novit curia*

Le decisioni dei giudici americani esaminate nel presente lavoro spaziano da alcune posizioni di assoluta chiusura ad altre assai permissive rispetto all'impiego dei *risk assessment tools* nel *sentencing*.

La *ratio decidendi* delle sentenze, come visto, ruota intorno al riconoscimento o meno di un abuso di discrezionalità del giudice nel momento in cui deve determinare l'ammontare della pena anche sulla base dei risultati prodotti dal *tool* algoritmico.

La riflessione ha mosso dalle argomentazioni assai radicali esposte in Indiana nei casi *Cardwell* e *Rhodes*, decisioni in cui si ostacola con forza *qualunque* utilizzo di algoritmi nel *sentencing* (si cfr. *supra*, § 4). In queste sentenze si scrive chiaramente che il ricorso a formule algoritmiche priva il giudice della possibilità di

¹²⁰ Parere del Comitato economico e sociale europeo sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, 22 dicembre 2021.

giudicare in autonomia tutti i fattori rilevanti per determinare la pena, inaridisce il suo giudizio discrezionale e lo deresponsabilizza.

In una prospettiva quasi del tutto opposta si posizionano invece le decisioni assunte nei casi *Malenchik e J.S.* (si cfr. *supra*, § 5), nelle quali si argomenta che proprio l'impiego di algoritmi di previsione del rischio di recidiva, che rappresentano uno dei tanti elementi da considerare nel procedimento di *sentencing*, permetterebbe di rinvigorire e potenziare (*enhance*) la valutazione autonoma del giudice attraverso la considerazione di un elemento aggiuntivo in grado di indirizzare la decisione verso il trattamento più adeguato per l'imputato.

Le motivazioni spese da ultimo dai giudici dell'Indiana diventano presto il perno su cui ruoterà la giurisprudenza successiva di altri stati. Sintetizzando, l'abuso di discrezionalità non si configura in quanto, ancorché esistano dubbi sull'affidabilità degli strumenti utilizzati, sulla loro provenienza privata e sulla loro conseguente inaccessibilità, essi non catturano né assorbono la decisione (*reliance*) ma costituiscono solo uno dei tanti elementi che il giudice può prendere in considerazione (*consider*) per assumere la decisione sulla pena. In questo senso non vi è *abuse of discretion* né quando il giudice trascura l'algoritmo (si cfr. *Samsa, supra*, § 5) né quando la pena è determinata assegnando un peso rilevante al risultato statistico (si cfr. *Loomis e People v. Younglove, supra*, § 5).

Sulla base di siffatte decisioni alcuni giudici hanno poi segnalato come la distinzione tra *reliance* e *consideration* appare in realtà ingannevole e nasconde la propensione mostrata dai giudici ad assegnare un peso decisivo all'algoritmo predittivo (si cfr. *supra*, § 6, spec. *Gordon* e la *concurring opinion* della giudice Roggensack in *Loomis*).

La tendenza ad affidare la decisione al risultato automatizzato comporta quindi che una parte della decisione non rimane affidata all'amministrazione pubblica ma viene esternalizzata a soggetti privati che progettano e immettono nel mercato gli algoritmi (si cfr. *supra*, § 7).

Tale fenomeno si aggrava ulteriormente se si considerano le caratteristiche degli algoritmi più evoluti, il cui funzionamento, come segnalato in altre sentenze (si cfr. *Gordon, Guise* e la *concurring opinion* della Giudice Cavanagh in *People v. Canedo*), molto spesso sfugge a chi è chiamato a utilizzarli. Si è visto inoltre che allo stato attuale in cui si trova la fisionomia degli algoritmi utilizzati e la preparazione dei giudici sembrerebbe arduo riconoscere nella motivazione un rimedio *ex se* efficace per ricondurre la discrezionalità su binari coerenti con la funzione del giudice (si cfr. *supra*, § 8).

Infine, nemmeno i primi passi compiuti dai legislatori europei per regolare il fenomeno – in particolare la progettazione pubblica e interdisciplinare degli algoritmi impiegati nel processo – parrebbero consentire di eliminare del tutto le preoccupazioni in ordine all'esercizio abusivo della discrezionalità (si v. *supra*, § 9).

Ora, ciò che sembra accomunare tutte le decisioni esaminate è l'assenza di una cornice che possa racchiudere il quadro complesso in cui si è sviluppata la giurisprudenza americana. Come già più volte accennato, tale cornice deve essere ricercata in una progressiva e imprescindibile formazione tecnica del giudice sugli algoritmi predittivi e la conseguente capacità di comprendere come l'algoritmo conduce propri calcoli, come da determinati dati viene prodotto l'indice di recidiva, come funziona il *machine learning*. Ciò in quanto non è possibile convincersi della utilità di un continuo sviluppo dell'intelligenza artificiale nell'amministrazione della giustizia se coloro i quali sono chiamati a farne applicazione non ne hanno il pieno controllo.

La necessità di una formazione adeguata del giudice è del resto avvertita anche dalle stesse Corti americane. In *Loomis*, l'imputato nel procedimento dinanzi alla *trial court* aveva chiamato a testimoniare un esperto criminologo, il quale aveva opportunamente riferito che le «*[s]entencing courts have very little information about how a COMPAS assessment analyzes the risks*»¹²¹, specificando che «*[t]he Court does not know how the COMPAS compares that individual's history with the population that it's comparing them with. The Court doesn't even know whether that population is a Wisconsin population, a New York population, a California population. . . . There's all kinds of information that the court doesn't have, and what we're doing is we're mis-informing the court when we put these graphs in front of them and let them use it for sentences*»¹²².

A differenza della Corte distrettuale, che non accoglie gli argomenti del perito, la Corte suprema pare invece condividere le perplessità sollevate rispetto al *deficit* di preparazione del giudice e, citando nella motivazione le parole tratte da uno studio specialistico in cui si sosteneva che «*unless criminal justice system actors are made fully aware of the limits of the tools they are being asked to implement, they are likely to misuse them*»¹²³, la Corte afferma di prestare attenzione al suddetto monito, imponendo alcune cautele per l'utilizzo dei *risk assessment tools*.¹²⁴

¹²¹ *Loomis*, § 27.

¹²² *Ibidem*.

¹²³ *Loomis*, § 43.

¹²⁴ Scrivono i giudici: «*We heed this admonition*», § 44. A ulteriore conferma di siffatta consapevolezza, i giudici supremi ritengono che l'utilizzo di COMPAS debba essere subordinato a un serie di cautele di cui i giudici devono tenere conto. Nella specie: 1. «*The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined*». 2. «*Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders, not a particular high-risk individual*». 3. «*Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism*». 4. «*A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations*». 5. «*COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole*», § 100. Soprattutto da quest'ultima cautela – vale a dire la considerazione che COMPAS non è stato progettato per venire impiegato dal giudice nel *sentencing* – pare evidenziare efficacemente il cortocircuito che si manifesta quando si fornisce al giudice uno strumento siffatto, progettato per altri fini e consegnato *sic et simpliciter* all'autorità giudicante che non possiede alcuna preparazione per poterlo utilizzare in modo adeguato.

Le cautele imposte dai giudici del *Wisconsin* non appaiono tuttavia sufficienti a dissipare le preoccupazioni rispetto all'impiego di tali sistemi, come peraltro sembra evincersi, ancora più chiaramente, da una *concurring opinion* della giudice Abrahamson nella stessa sentenza *Loomis*, nella quale afferma che «*this court's lack of understanding of COMPAS was a significant problem in the instant cases*», che rischia di infirmare il «*basic requirement that a circuit court explain its exercise of discretion at sentencing*»¹²⁵.

Ancora, in una relazione prodotta in giudizio nel caso *Samsa* (si v. *supra*, § 5), l'imputato notava opportunamente che: «*[w]ithout training on the COMPAS assessment, it would be easy to misinterpret the results, which are often displayed in bar chart form*»¹²⁶.

Dunque, senza un'adeguata preparazione sul funzionamento dell'algoritmo, è verosimile che il risultato venga male inteso o interpretato, impedendo al giudice di motivare adeguatamente la sua decisione discrezionale. Siffatte osservazioni, che evidenziano l'incapacità di comprendere gli algoritmi impiegati nel *sentencing*, conducono a domandarsi quale sia lo spazio in cui gli algoritmi predittivi potrebbero venire impiegati nel sistema penale americano.

Se non ci si vuole accontentare di una soluzione che neghi *sic et simpliciter* l'impiego dei *tools* (che sarebbe probabilmente antistorica, vista l'irrefrenabile evoluzione tecnologica), scavando sul fondo delle parole dei giudici – consci ormai, arrivati sino a qui, delle complessità che attengono all'impatto dei *tools* sulla giurisdizione – si intravedono alcuni spazi di riflessione, seppure ancora parziale e incompleta, legati all'aggiornamento professionale del giudice e al rinnovamento anche culturale del suo operare.

La preventiva e profonda consapevolezza del giudice sul funzionamento dei *tools* getterebbe, infatti, nuova luce su tutte le sentenze qui prese in esame: in questo senso le decisioni di assoluta chiusura all'impiego di algoritmi, che muovevano proprio dal disagio mostrato dai giudici verso strumenti complessi e oscuri, la cui formula *mindless* sarebbe «inevitabilmente iniqua», apparirebbero poco condivisibili, in quanto l'impiego di strumenti algoritmici in grado di produrre ed elaborare dei dati difficilmente acquisibili dall'essere umano, purché mantenuti nel circuito pubblico e controllati da un giudice in grado di comprenderne le peculiarità e le eventuali storture, fornirebbero allo stesso decidente un elemento di prova utile per una decisione più accurata e razionale.

Di contro, le sentenze che qualificano i *risk assessment tools* come strumenti in grado di migliorare la decisione discrezionale del giudice e che distinguono frettolosamente la legittimità della *consideration* dall'illegittimità della *reliance* acquisirebbero un significato diverso. Detta distinzione, se oggi, come visto, non trova adeguato fondamento, potrebbe consentire in futuro al giudice, qualora accompagnata da una adeguata preparazione professionale, di motivare sulla rilevanza nella decisione dell'algoritmo stesso e

¹²⁵ *Loomis*, § 141.

¹²⁶ § 10.

sulle ragioni per le quali in un caso ha deciso di discostarsi dal risultato prodotto, valorizzando altri elementi della causa, in un altro decidendo di tenerlo, invece, in maggiore considerazione.

Infine, il pieno controllo degli strumenti algoritmici, oltre alle altre cautele afferenti alla trasparenza degli stessi e alla loro progettazione pubblica, rallenterebbe il fenomeno della esternalizzazione e riporterebbe il giudizio discrezionale nelle mani del giudice, il quale si assumerebbe la piena responsabilità della decisione finale.

Da tale angolo prospettico l'*abuse of discretion* di cui abbiamo discusso assume una nuova fisionomia, potendosi configurare un abuso tutte le volte in cui la discrezionalità viene esercitata senza una previa consapevolezza degli elementi su cui la stessa discrezionalità dovrebbe fondarsi.

La rivoluzione dell'intelligenza artificiale, quindi, se vuole insinuarsi tra le pieghe del processo, non passa solo da una sofisticazione della tecnica, dalla ricerca di miglioramento dei sistemi artificiali (ovviamente necessaria), ma, soprattutto, da una rivoluzione culturale dei giudici, non dissimile da quella che hanno dovuto compiere con l'avvento del costituzionalismo¹²⁷ e del processo di integrazione europea¹²⁸.

Ciò su cui si dovrà, nei prossimi tempi, lavorare non è tanto la ricerca di un'intelligenza artificiale più coerente con i principi che reggono lo stato democratico costituzionale – che rimane imprescindibile – quanto piuttosto la formazione dei giudici che con questa intelligenza artificiale dovranno fare i conti e stanno già facendo i conti¹²⁹. Se ora i giudici appaiono ancora riluttanti a “decodificare” l'intelligenza artificiale¹³⁰, con una formazione adeguata avranno più strumenti per mantenere il controllo della decisione e – in fondo – per riappropriarsi del potere discrezionale di cui stanno venendo piano piano spogliati.

¹²⁷ Si cfr., sul nuovo ruolo del giudice con l'avvento del costituzionalismo, tra i tanti, P. GROSSI, *Ritorno al diritto*, Laterza, Roma-Bari, 2015.

¹²⁸ Sono note le difficoltà che ha dovuto affrontare il giudice nell'interpretazione e applicazione della legge quando ha dovuto iniziare a fare i conti sia con il sistema del Consiglio d'Europa (e quindi con la giurisprudenza della Corte EDU), sia con le fonti dell'Unione Europea (e la giurisprudenza della Corte di Giustizia). Su tali tematiche, soprattutto con riguardo alle problematiche afferenti il rapporto tra giudici domestici e CEDU, si cfr., B. RANDAZZO, *Interpretazione delle sentenze della Corte europea dei diritti ai fini dell'esecuzione (giudiziaria) e interpretazione della sua giurisprudenza ai fini dell'applicazione della CEDU*, in *Rivista AIC*, 2/2015. Già prima si cfr. B. RANDAZZO, *Giudici comuni e Corte europea dei diritti*, in *Atti del Seminario del “Gruppo di Pisa”, La Corte costituzionale e le Corti d'Europa*, (Catanzaro, 31 maggio – 1° giugno 2002), Giappichelli, Torino 2003, pp. 217-288. Sul mutamento ruolo del giudice in generale si cfr., *ex multis*, R. ROMBOLI, *Il ruolo del giudice in rapporto all'evoluzione del sistema delle fonti ed alla disciplina dell'ordinamento giudiziario*, in *www.associazionedeicostituzionalisti.it*, 2005.

¹²⁹ Nella prospettiva della tutela *by education*, proposta da A. SIMONCINI in *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, cit., 88. In realtà l'A. utilizza tale espressione per sostenere la necessità di una formazione giuridica per i tecnici che progettano l'algoritmo. Invece, nella prospettiva qui proposta, viene impiegata per indicare la necessità di una formazione tecnica dei giudici.

¹³⁰ In dottrina si è fatto notare che «*the court was disturbed at the lack of evidence it received as to exactly how this calculation had been done but, interestingly, did not seem interested to find out more but rather to exclude it from influence. It seems quite likely that courts will be reluctant to become activists about disclosures of source code, let alone algorithmic training sets and models, until they feel more confident of their ability to comprehend and use such evidence - which may take some time*» cfr. L. EDWARDS - M. VEALE, *Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?*, in *IEEE Security & Privacy*, 16(3), 2018, p. 12.

In altre parole, solo tramite l'accorto utilizzo degli strumenti di *risk assessment* – che non può prescindere da un rinnovamento professionale del giudice – «*the assessment tool scores may, and if possible should, be considered to supplement and enhance a judge's evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant*»¹³¹.

In definitiva, rivolgendo lo sguardo anche al nostro ordinamento, l'ingresso di sistemi di intelligenza artificiale nel processo penale deve essere guardato con grande cautela se esso non andrà di pari passo con (o meglio, se non verrà anticipato da) un'adeguata formazione del nuovo “giudice tecnologico”, che sia in grado di mantenere presso di sé la decisione discrezionale.

Si staglia dunque all'orizzonte un nuovo principio, che dovrà fondare l'eventuale decisione giudiziale “promiscua” (assunta cioè da un giudice con l'ausilio dell'algoritmo) e che si può mutuare, con qualche inevitabile adattamento, dall'aforismo *iura novit curia*. Come si presuppone che il giudice sia a conoscenza delle norme e abbia l'obbligo di farne applicazione anche in mancanza di una corretta individuazione delle parti, così questi dovrà conoscere l'algoritmo per poter individuare le storture e per poter eventualmente decidere di discostarsene: *algorithmia novit curia*.

Solo assumendo come stella polare questo nuovo principio sarà immaginabile un futuro in cui la giustizia, razionalizzata e migliorata dall'impiego dell'intelligenza artificiale, rimarrà appannaggio di una coscienza «viva, sensibile, vigilante, umana»¹³² dell'autorità pubblica – il giudice – e non di un soggetto privato, *alienus* alle più basilari esigenze di giustizia presenti in qualunque ordinamento democratico e costituzionale.

¹³¹ Sono parole tratte dalla sentenza *Malenchik*, si v. *supra*, § 5.

¹³² P. CALAMANDREI, *Processo e Democrazia*, in M. CAPPELLETTI (a cura di), *Opere giuridiche*, vol. I, Morano, 1965, p. 647. L'A. scriveva al riguardo che «il giudice non è un meccanismo, non è una macchina calcolatrice [...] Ridurre la funzione del giudice a un puro sillogizzare vuol dire impoverirla, inaridirla, disseccarla». Sempre in tempi risalenti il giudice americano Frankfurter, nella decisione *Rochin v. California* 342 U.S. 165, 171 del 1952 scriveva nella motivazione: «*To believe that this judicial exercise of judgment could be avoided by freezing “due process of law [...] is to suggest that the most important aspect of constitutional adjudication is a function for inanimate machines and not for judges [...]. Even cybernetics has not yet made that haughty claim*».