

Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?

I lavori in corso alla Commissione europea
per l'efficacia della giustizia (Cepej) del Consiglio d'Europa

di *Clementina Barbaro*

L'uso dell'intelligenza artificiale come strumento di supporto del lavoro degli operatori del diritto e dei tribunali è ancora un fenomeno embrionale in Europa. Tra le diverse soluzioni proposte, suscitano un interesse crescente quelle riguardanti l'analisi e il trattamento della giurisprudenza per ridurre l'alea del giudizio e garantire una maggiore prevedibilità delle decisioni giudiziarie. È necessario, tuttavia, interrogarsi sull'effettività di queste applicazioni, impropriamente qualificate di "giustizia predittiva", alla luce delle caratteristiche attuali dell'intelligenza artificiale e, in particolare, del *machine learning/apprendimento automatico*. Il loro impatto sulla professione del magistrato, sulla produzione della giurisprudenza e sulle garanzie dell'equo processo merita inoltre di essere attentamente esaminato.

La Commissione europea sull'efficacia della giustizia (Cepej) ha condotto uno studio approfondito su questi temi e redige attualmente una «Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari», adottata alla sessione plenaria della Cepej del 3-4 dicembre 2018. Primo strumento europeo in materia, la Carta etica enuncia principi sostanziali e metodologici applicabili all'analisi e al trattamento delle decisioni giudiziarie e intende essere un punto di riferimento per l'attività di soggetti privati e pubblici attivi in questo settore, tanto per quanto riguarda lo sviluppo concreto di applicazioni di intelligenza artificiale quanto per l'elaborazione di politiche pubbliche riguardanti l'integrazione di tali applicazioni nel sistema giudiziario.

La Commissione europea sull'efficacia della giustizia (Cepej) del Consiglio d'Europa si interessa da diversi anni all'impatto delle tecnologie dell'informazione e della comunicazione sui sistemi giudiziari europei. Nel 2016, la Cepej ha adottato uno studio approfondito sull'uso di queste tecnologie nei tribunali europei¹ e delle Linee direttrici sulla "cybergiustizia"². In continuità con il lavoro svolto e alla luce di importanti sviluppi in alcuni sistemi giudiziari europei – in particolare, il progressivo diffondersi di applicazioni fondate sull'intelligenza artificiale (Ia) a fini di soste-

gno dell'attività giudiziaria –, la Cepej ha intrapreso l'elaborazione di una «Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari».

La Carta etica è stata adottata, in prima battuta, dal Gruppo di lavoro sulla qualità della giustizia, composto da sei membri (magistrati, professori universitari, funzionari ministeriali) nominati dal *Bureau* della Cepej su proposta dei rispettivi Governi, ed è stata formalmente adottata dalla Cepej durante la sessione plenaria del 3-4 dicembre³. I suoi contenuti saranno dettagliati ulteriormente in questo articolo,

1. Cepej, Studio n. 24, *Rapport thématique: l'utilisation des technologies de l'information par les tribunaux en Europe*, 2016 (dati del 2014).

2. Cepej (2016) 13, <https://edoc.coe.int/fr/efficacite-de-la-justice/7498-lignes-directrices-sur-la-conduite-du-changement-vers-la-cyberjustice-bilan-des-dispositifs-deployes-et-synthese-de-bonnes-pratiques.html>.

3. Per il testo della «Carta etica europea sull'uso dell'intelligenza artificiale (Intelligenza artificiale) nei sistemi giudiziari e in ambiti connessi» (in inglese e francese) e per un primo commento, cfr. C. Barbaro, *Cepej, adottata la prima Carta etica europea sull'uso dell'in-*

ma si evidenzia fin d'ora che la Carta è il primo strumento europeo a enunciare i principi fondamentali che dovrebbero guidare l'applicazione dell'Intelligenza artificiale nei sistemi giudiziari, particolarmente quando le soluzioni di Intelligenza artificiale si fondano sul trattamento delle decisioni giudiziarie. Alla formulazione dei principi seguiranno uno studio approfondito di alcune applicazioni di Intelligenza artificiale in corso di sperimentazione in diversi Stati europei e delle raccomandazioni sull'uso di tali applicazioni, un glossario e una *checklist* di autovalutazione, per verificare il livello di adesione ai principi summenzionati.

Questo articolo illustra, in maniera succinta, le linee portanti della riflessione del Gruppo di lavoro sulla qualità della giustizia su questo tema⁴. A questo proposito, si nota che le applicazioni qualificate di "giustizia predittiva" sono state oggetto di particolare attenzione da parte del Gruppo: da un lato, perché vivaci dibattiti sono in corso in alcuni Paesi europei con riferimento al loro ambito di applicazione e ai loro effetti sul lavoro dei magistrati; dall'altro, perché esse sono particolarmente rappresentative delle caratteristiche di funzionamento dei sistemi di Intelligenza artificiale attuali. Queste due questioni sono, in realtà, interconnesse: una riflessione sull'uso di queste tecnologie applicate alla giurisprudenza (par. 2) non può, in effetti, prescindere da un'analisi obiettiva delle loro potenzialità e dei loro limiti di funzionamento (par. 1). Inoltre, la definizione di principi etici è essenziale al fine di garantire che gli strumenti di Intelligenza artificiale, in particolare quelli che si basano sul trattamento delle decisioni giudiziarie, incluse quelle di "giustizia predittiva", soddisfino un certo numero di criteri sostanziali e metodologici (par. 3).

1. Meccanismi alla base degli strumenti di "giustizia predittiva" e loro limiti intrinseci

Si rileva, in primo luogo, che con l'espressione "intelligenza artificiale" si designa un insieme di scienze, teorie e tecniche il cui obiettivo ultimo è quello di permettere a una macchina di risolvere problemi,

comprendere e interagire con la realtà come farebbe un essere umano. Gli specialisti distinguono tra intelligenze artificiali "forti" (in grado di contestualizzare e risolvere in modo autonomo problemi specialistici molto diversi tra loro) e intelligenze artificiali "deboli" o "moderate" (in grado di fornire prestazioni specifiche qualitativamente equivalenti e quantitativamente superiori a quelle umane). Le intelligenze artificiali utilizzate per il trattamento e l'analisi della giurisprudenza, su cui si fondano le applicazioni di "giustizia predittiva", appartengono a questa seconda categoria e si basano in particolare su due tecniche:

- il *natural language processing* (trattamento del linguaggio naturale): ci si riferisce qui, in estrema sintesi, al trattamento informatico del linguaggio umano⁵;
- il *machine learning* (l'apprendimento automatico); questa tecnica permette di costruire uno o più modelli matematici una volta identificate le correlazioni esistenti tra grandi masse di dati. A differenza dei "sistemi esperti" degli anni '90, che cercavano di modellizzare delle regole di gestione fondate sul ragionamento umano, e di applicarle ai dati in entrata, la macchina ricerca in modo autonomo (con o senza la supervisione di un essere umano) le correlazioni e deduce delle regole dall'enorme quantità di dati forniti. Applicata alle decisioni giudiziarie, essa individua le correlazioni tra i diversi parametri di decisioni rese in passato su un certo argomento (in materia di divorzio, ad esempio, può trattarsi della durata del matrimonio, del reddito dei coniugi, il verificarsi di un adulterio, etc.) e ne deduce uno o più modelli matematici, su cui si fonda per "prevedere" gli elementi di una futura decisione giudiziaria (ad esempio, l'ammontare dell'assegno di mantenimento).

Le modalità di funzionamento di cui sopra inducono una riflessione sull'uso dell'aggettivo "predittivo", che è frequentemente utilizzato nell'ambito della matematica, dove per esempio si parla di "equazione predittiva". Peraltro, se la facciamo risalire alle radici latine, la predizione è l'atto di annunciare (*prae*, prima - *dicere*, dire) eventi futuri con un'ispirazione

telligenza artificiale (AI) nei sistemi giudiziari, in *Questione giustizia on line* del 7/12/2018, www.questionegiustizia.it/articolo/cepej-a-dottata-la-prima-carta-etica-europea-sull-uso-dell-intelligenza-artificiale-ai-nei-sistemi-giudiziari_07-12-2018.php.

4. Vorrei ringraziare il Dottor Yannick Meneceur, magistrato, consigliere di intelligenza artificiale al Consiglio d'Europa. Yannick Meneceur ha contribuito ai lavori del Gruppo ed è autore di numerose pubblicazioni sulla giustizia predittiva, tra cui *Le numérique, levier essentiel d'une meilleure efficacité et qualité de la justice en Europe*, in *Enjeux numériques*, Les Annales des mines, Parigi, n. 3/2018, pp. 11-15 (disponibile online: www.anales.org/enjeux-numeriques/2018/en-2018-03/EN-2018-09-3.pdf).

5. Applicazioni di uso comune che utilizzano questa tecnica sono, ad esempio, le assistenti vocali («Siri» di Apple o «Alexa» di Amazon, per citarne alcune tra le più conosciute).

quasi soprannaturale. La previsione, invece, è il risultato dell'osservazione (*videre*, vedere) di un insieme di dati, al fine di prevedere una situazione futura. Quindi, ancora prima di discutere i meriti di questi sistemi che cercano di anticipare le decisioni giudiziarie, sembra imperativo qualificarli per quello che fanno realmente (della previsione e non della predizione), mettendo da parte un linguaggio fuorviante.

Poste queste premesse, per meglio comprendere le caratteristiche delle applicazioni di apprendimento automatico applicato alla giurisprudenza, la Ceppej ha fatto riferimento a uno studio condotto dalla *University College of London* (Ucl), che si è occupato di 584 decisioni di apprendimento automatico della Corte europea dei diritti dell'uomo⁶. Questo studio è stato scelto per diversi motivi: è l'unico studio che si è occupato della giurisprudenza della Corte di Strasburgo e, soprattutto, è pubblico e ben documentato. I risultati sono stupefacenti: con l'utilizzo di tecniche di apprendimento automatico, la macchina riproduce le decisioni umane con una percentuale di successo del 79%.

Questo risultato è, in realtà, tanto sorprendente quanto la meccanica utilizzata. La macchina è stata programmata su diversi elementi costitutivi delle sentenze della Corte: i *fatti*, l'*applicazione della Convenzione europea dei diritti dell'uomo* e il *dispositivo* (binario: violazione o non violazione). La frequenza della presenza di gruppi lessicali coerenti è stata, poi, registrata in una banca dati e confrontata con la violazione (o meno) della Convenzione. In sintesi, le parole che potrebbero portare a una violazione (o non violazione) della Convenzione sono state ponderate in base alla loro presenza predominante nei diversi tipi di decisioni esaminate. Così, secondo lo studio, i termini «condizioni di detenzione», «obblighi dello Stato», «esecuzione delle decisioni» sembravano avere maggiori probabilità di essere trovati in una decisione di violazione. Al contrario, la «mancanza di prove», i «diritti di proprietà», la «precedente violazione degli articoli 2 e 11» sembravano avere maggiori probabilità di essere trovati nelle decisioni di non violazione.

La macchina Ucl *non ha dunque cercato di simulare un ragionamento giuridico*; ha, semplicemente, provveduto al trattamento statistico dei dati raccolti e calcolato delle probabilità. Lo stesso lavoro, svolto sull'applicazione della Convenzione, ha portato a una riduzione del tasso di successo della riproduzione

delle decisioni al 62%. Questo risultato, più che soddisfacente a prima vista, deve tuttavia essere riconsiderato alla luce del fatto che chiunque ha una possibilità di “predire” l'esito di una decisione affermando a caso “violazione” o “non violazione” (50%).

Si potrebbe obiettare che la tecnica utilizzata è solo una tra le tante, e che altre soluzioni sono testate da ricercatori ed esperti informatici per superare questi limiti. Tuttavia, le questioni di fondo che queste nuove tecnologie cercano di trattare sono le seguenti: possiamo modellare matematicamente la legge e la sua applicazione? L'approccio “inferenziale” del *machine learning* può rivelare delle regole di applicazione del diritto? In altre parole, questo tipo di tecnica può illustrare delle *tendenze nella giurisprudenza* o addirittura rivelare *pregiudizi nel comportamento dei giudici*, come hanno sostenuto alcuni articoli a supporto della “giustizia predittiva” pubblicati in Francia?⁷

Per rispondere a questo quesito, e valutare altresì la pertinenza delle soluzioni proposte, è utile ricordare quali siano le caratteristiche principali di una decisione giudiziaria. Una decisione giudiziaria è, essenzialmente, la selezione dei fatti rilevanti da parte del giudice, l'interpretazione e l'applicazione delle norme giuridiche a tali fatti. L'applicazione del diritto ai fatti è spesso un'operazione complessa: ci possono essere, per esempio, una moltitudine di fonti che non si integrano in modo perfettamente coerente, caratteristica che il teorico del diritto Herbert L.A. Hart ha qualificato di «tessuto aperto del diritto»⁸. Inoltre, il sillogismo giudiziario sembra essere più un modo di presentare il ragionamento giuridico *a posteriori* che la traduzione logica di tutte le tappe del ragionamento effettuato dal giudice. In effetti, tale ragionamento può essere punteggiato da una moltitudine di scelte discrezionali che non sono riconoscibili *a priori*, come il fatto di selezionare e di trattare solo alcuni degli elementi addotti dalle parti.

Una decisione giudiziaria costituisce, dunque, una materia prima imperfetta per consentire un'identificazione rigorosa di tutti i fattori causali di una decisione, creare dei modelli di ragionamento del giudice e raggiungere dei risultati “predittivi” affidabili. Per le stesse ragioni, non è possibile affermare che il *machine learning* permetta di spiegare il comportamento del giudice. Il fatto che un giudice renda statisticamente delle decisioni di affidamento dei figli

6. N. Aletras - D. Tsarapatsanis - D. Preotiuc-Pietro - V. Lampos, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, in *PeerJ computer science*, 24 ottobre 2016, <https://peerj.com/articles/cs-93/>.

7. M. Benesty, *L'impartialité des juges mise à mal par l'intelligence artificielle*, in *Village de la justice*, 24 marzo 2016, www.village-justice.com/articles/impartialite-certains-juges-mise,21760.html.

8. H.L.A. Hart, *Il concetto di diritto*, Einaudi, Torino, 2002 (ed. or.: *The Concept of Law*, Clarendon Press, Oxford, 1961).

a favore della madre non riflette necessariamente un pregiudizio nei confronti del padre, ma piuttosto l'esistenza di fattori psicosociali, economici o addirittura culturali specifici della giurisdizione in cui esercita, come la presenza o meno di servizi sociali, il fatto che il bambino vada a scuola oppure no, l'orario di lavoro di ciascun genitore, o anche semplicemente la mancanza di interesse da parte di uno dei due genitori a farsi carico del figlio piccolo.

Il rischio che si pone concretamente è che le correlazioni operate dal *machine learning* siano in realtà poco significative e che un lavoro analitico ulteriore sia necessario per isolare i veri elementi causali di una decisione.

Per le ragioni summenzionate, le prestazioni dell'Intelligenza artificiale applicata alla giurisprudenza si rivelano dunque comparativamente più deboli e imperfette che in altri campi, in cui i risultati raggiunti sono eccezionali. Citiamo qui, per il clamore mediatico che ha suscitato, la vittoria dell'applicazione di Intelligenza artificiale «Alpha Go» sul campione mondiale (umano) del gioco cinese molto complesso del Go nel 2012, e la successiva vittoria di «Alpha Go Zero» (versione ulteriore potenziata) su «Alpha Go» nel 2017⁹. Ma perché l'Intelligenza artificiale funziona in maniera estremamente efficace con il gioco del Go e non con il diritto?

Per rispondere a questa domanda, bisogna tenere conto di un limite importante dell'Intelligenza artificiale, che è l'incapacità di adattare il suo funzionamento al di fuori del proprio modello. Il mondo del gioco del Go è definito e definibile per una macchina: pietre bianche o nere, 19 linee su 19, combinazioni esponenziali di movimento, ma senza che sia possibile effettuare un completo cambiamento di paradigma. Gli scienziati sono in grado di modellare questo tipo di mondo utilizzando la matematica discreta, che si occupa di insiemi numerabili o finiti, dove si possono effettuare milioni di prove per gestire le milioni di combinazioni possibili. Bisogna ammettere che la matematica ci ha insegnato che alcuni eventi aleato-

ri possono essere modellati, come nella *Macchina di Galton*, dove un gran numero di palline cadono verticalmente attraverso file di chiodi prima di depositarsi sul fondo, formando delle pile. Nonostante le palline si muovano a caso a destra o a sinistra, al termine dell'esperimento, le altezze di queste pile assumono approssimativamente la forma di una curva a campana. Ma cosa succederebbe se ci fosse un cambiamento totale di paradigma in questo modello?

Il mondo reale, che non è finito come quello del Go, ma complesso e continuo in logica matematica, rappresenta di per sé un limite importante per l'Intelligenza artificiale: come menzionato in precedenza, quest'ultima può tentare di identificare regole implicite, ma non è in grado di isolare significativamente tutti i fattori causali e non può cambiare radicalmente i modelli per cui è progettata. Il *machine learning* e la *data science*¹⁰ permettono di realizzare delle prestazioni superiori a quelle dell'essere umano in certi ambiti (per esempio, l'analisi quantitativa di dati), ma non forniscono una soluzione per rivelare indicatori qualitativi significativi e costruire modelli di proiezione affidabili nel campo (molto imperfetto) della giurisprudenza. Inoltre, la società si evolve, le regole del diritto si evolvono e i giudici stessi si evolvono. Le decisioni giudiziarie non sono come palline cadute su una *Macchina di Galton*: il ragionamento giuridico non è solo la distribuzione casuale di opzioni binarie. Come prevedere, sulla base del resoconto fatto dalle parti, quali fatti saranno considerati rilevanti dal giudice? Come ponderare elementi di apprezzamento come l'equità?

In conclusione, lo stato di sviluppo delle tecniche di apprendimento automatico attuali non consente ancora di ottenere risultati affidabili per quanto riguarda la "previsione" delle decisioni giudiziarie. Gli esperimenti condotti in Francia, su scala limitata, in due corti d'appello hanno, d'altronde, rilevato "la mancanza di valore aggiunto" per il lavoro dei giudici di un'applicazione "predittiva" fornita da una società privata¹¹.

9. «Alpha Go Zero», versione potenziata di «Alpha Go», è riuscita a battere quest'ultima in tre giorni dopo essere stata formata unicamente sulle regole del gioco e non sulle partite precedenti effettivamente giocate.

10. Insieme di scienze (matematica, statistica, probabilità, informatica, visualizzazione dei dati) che mirano a estrarre conoscenze da un insieme eterogeneo di dati.

11. Su iniziativa del Ministero della giustizia, le Corti d'appello di Rennes e Douai hanno concordato di testare su vari casi d'appello, nella primavera del 2017, una versione di un *software* "predittivo" di analisi delle decisioni in materia civile e commerciale di tutte le corti d'appello francesi.

Anche se dati giurisprudenziali interni ed esaustivi erano già disponibili gratuitamente da molti anni (banca dati Jurica) e messi appositamente a disposizione della casa editrice dal Ministero, ai due tribunali è stato offerto di valutare il valore di un'analisi quantificata, presentata come innovativa, delle somme stanziati dai tribunali, nonché una classificazione geografica delle differenze riscontrate per domande e processi simili.

L'ambizione dichiarata del *software* era, quindi, quella di creare uno strumento decisionale per ridurre, se necessario, l'eccessiva variabilità delle decisioni giudiziarie, in nome del principio di uguaglianza dei cittadini di fronte alla legge. Il ritorno della sperimentazione, contraddittoriamente dibattuto tra le due Corti d'appello, il Ministero della giustizia e il *legaltech* all'origine del prodotto, ha purtroppo

2. L'intelligenza artificiale applicata alle decisioni giudiziarie

Nonostante questi limiti, sarebbe un grave errore non sfruttare le grandi potenzialità dell'Intelligenza artificiale per sostenere e migliorare l'efficacia del lavoro giudiziario. La "giustizia predittiva" rappresenta una delle tante applicazioni di Intelligenza artificiale: ve ne sono altre, molto interessanti, in fase di sperimentazione in alcuni Paesi europei – per esempio, l'uso dell'Intelligenza artificiale per la lettura rapida, la classificazione e l'attribuzione di atti, ricorsi e documenti alle sezioni di cancelleria pertinenti, oggetto di *test* in Austria; o l'uso di *chatbot* (letteralmente: conversazione con un robot) per orientare il cittadino verso una soluzione di risoluzione alternativa della controversia, come in Lettonia.

Con riferimento in particolare alle applicazioni che si fondano sul trattamento delle decisioni dei tribunali a fini di ricerca giuridica, si rileva che esistono già, allo stato attuale, motori di ricerca "intelligenti", che permettono di analizzare fonti e materiali diversi (leggi, regolamenti, giurisprudenza, dottrina). L'Intelligenza artificiale può permettere di avere un accesso all'informazione non solo più largo e diversificato, ma anche più interattivo: i giudici, gli avvocati e altri professionisti potrebbero "navigare" tra le diverse informazioni e trovare più rapidamente le soluzioni ricercate. L'Intelligenza artificiale potrebbe essere anche usata per proporre modelli o estratti di una decisione che siano in correlazione coi risultati della ricerca, rendendo così più rapida la redazione da parte del giudice.

I metodi di calcolo delle indennità in materia civile (ad esempio, il risarcimento per danni fisici, l'indennità di licenziamento) potrebbero essere notevolmente migliorati, con un accesso a una più ampia base di decisioni di quella attuale, a condizione di adottare una serie di misure precauzionali per la progettazione e l'uso di cui si dirà più diffusamente al par. 3. Ci riferiamo qui al cosiddetto effetto "performativo" o di autorealizzazione, cioè al rischio che un sistema produca progressivamente gli stessi risultati in uscita influenzando i produttori di informazioni in entrata (i magistrati). È questa, d'altronde, una delle

critiche principali che viene mossa agli strumenti di "giustizia predittiva", cioè il fatto che la predizione non abbia più un valore *indicativo*, ma quasi *prescrittivo* per il giudice. Quest'ultimo tenderebbe ad applicare la soluzione proposta dalla macchina senza esaminarla alla luce delle particolarità del caso oggetto di esame, alimentando così il sistema e facilitando la riproduzione meccanica di decisioni rese precedentemente. Questi sistemi condurrebbero dunque a una "cristallizzazione" della giurisprudenza e potrebbero limitare la sua evoluzione.

Sarebbe possibile utilizzare tali calcoli per indirizzare le parti in causa verso metodi alternativi di risoluzione delle controversie? A prima vista, sembrerebbe allettante predisporre dei meccanismi di trattamento più o meno automatizzati della giurisprudenza per limitare contenziosi civili ripetitivi e di basso valore. Gran Bretagna, Paesi Bassi e Lettonia sono esempi di Paesi che hanno già attuato, o stanno per attuare, questo tipo di soluzione. Ma su quale base sarà calcolata la compensazione? Con quale metodo? L'algoritmo è in grado di elaborare correttamente le informazioni? La proposta può essere discussa in modo contraddittorio con l'aiuto di una terza parte qualificata e certificata? L'accesso al giudice è sempre possibile? È essenziale che principi di trasparenza, di neutralità e lealtà verso i cittadini siano integrati nei meccanismi di trattamento in questione.

In materia penale, si assiste a una diffusione crescente degli strumenti di "polizia predittiva" (pensiamo, ad esempio, alla *no fly list*, una applicazione che raccoglie e analizza dati su potenziali terroristi con l'obiettivo di prevenire la commissione di atti criminali; o, ancora, alla "cartografia criminale predittiva", che utilizza l'Intelligenza artificiale per identificare luoghi di commissione di reati di natura seriale e per predisporre in maniera efficiente i servizi di polizia).

L'uso dell'Intelligenza artificiale nel quadro del processo penale è molto limitato in Europa. L'esperienza americana è, invece, significativa e fonte di riflessione anche per i Paesi europei che valutano l'introduzione di queste tecnologie nell'ambito del processo penale, come il Regno Unito¹².

Negli Stati Uniti, gli strumenti di determinazione del rischio (*risk assessment tools*) sono spesso uti-

obiettato la mancanza di valore aggiunto della versione testata del *software* per la riflessione e il lavoro decisionale dei giudici.

In particolare, sono stati rivelati errori di ragionamento del *software*, che hanno portato a risultati anomali o inappropriati, a causa della confusione tra semplici frequenze lessicali delle motivazioni giudiziarie e le causalità che erano state decisive nel ragionamento dei giudici in esame.

12. Nel Regno Unito, il *software* HART («*Harm Assessment Risk Tool*»), sviluppato in collaborazione con l'Università di Cambridge, è attualmente in fase di sperimentazione. Questa tecnologia, basata sull'apprendimento automatico, è stata programmata con le informazioni provenienti dagli archivi della polizia di Durham per oltre cinque anni, dal 2008 al 2012. Osservando le decisioni prese dai giudici durante questo periodo e la recidiva (o meno) di alcuni sospetti, la macchina dovrebbe essere in grado di valutare il rischio – basso, medio o alto – delle persone sottoposte a fermo di polizia, in particolare sulla base di una trentina di fattori, alcuni dei quali non legati al reato commesso (ad esempio, il luogo di residenza).

lizzati dal giudice, pur non essendo vincolanti ai fini della decisione, per determinare la probabilità di recidiva di una persona in stato di fermo, la sanzione (per esempio, la durata della pena detentiva) e l'eventuale imposizione di una misura alternativa alla detenzione. Queste probabilità (di livello alto, medio, o basso) sono misurate attraverso il trattamento statistico di diversi fattori di rischio riguardanti sia la persona in questione sia individui precedentemente fermati o giudicati, che hanno in seguito commesso un altro reato.

Negli Stati Uniti, l'Ong ProPublica ha rivelato gli effetti discriminatori di uno di questi strumenti di valutazione del rischio utilizzati in alcuni Stati federati. Si tratta specificatamente dell'algoritmo utilizzato nella *software Compas*, sviluppato da un'impresa privata, che serve a predire il rischio di recidiva. Alle popolazioni afro-americane è stato assegnato un tasso di rischio di recidiva entro due anni dalla sentenza che era il doppio di quello assegnato alle popolazioni bianche; al contrario, l'algoritmo riteneva che queste ultime fossero meno esposte a tale rischio¹³. Questo tipo di strumenti tende dunque a riprodurre o, addirittura, amplificare discriminazioni esistenti nel sistema penale, con effetti che si potrebbero qualificare *deterministi*, nel senso che inchiodano un individuo al destino di una comunità. Un risultato non sorprendente, se si tiene conto che l'algoritmo rielabora, in effetti, dati non neutri, riguardanti comunità che, più di altre, sono soggette a misure di polizia e a sanzioni penali.

Due aspetti meritano di essere particolarmente evidenziati. Il primo è l'opacità degli algoritmi utilizzati dalle aziende private e il rifiuto di queste ultime di rendere conto delle caratteristiche dei loro prodotti. Gli algoritmi, in effetti, sono protetti dal segreto industriale ed è difficile (salvo autorizzazione del giudice, come è avvenuto nell'ambito dell'inchiesta di ProPublica) avervi accesso. Posto che i dati di cui si nutre l'algoritmo provengono dalle autorità statali e che quest'ultimo è utilizzato nell'ambito dell'attività giudiziaria con incidenze significative sui diritti dei cittadini, non si può non rilevare un problema di trasparenza e di *accountability* degli attori privati. Il secondo aspetto che merita di essere sottolineato è la necessità di garantire il pieno rispetto del principio della parità delle armi tra accusa e difesa: la persona giudicata sulla base di un algoritmo dovrebbe avere la possibilità di esaminarne e contestarne la validità scientifica, nonché il peso attribuito alle sue diverse componenti.

Infine, delle preoccupazioni sussistono quanto all'impatto potenzialmente negativo che questi strumenti possono avere sull'imparzialità del magistrato: un giudice che decide contro la previsione di un algoritmo si assume un rischio e, di conseguenza, maggiori responsabilità. Non sembra irrealistico immaginare che i giudici possano essere riluttanti a farsi carico di questo onere, in particolare nei sistemi in cui il loro mandato non è permanente, ma soggetto al voto popolare, o in quelli in cui la loro responsabilità personale (disciplinare, civile o addirittura penale) può essere fatta valere, soprattutto se le garanzie previste dalla legge in materia disciplinare sono insufficienti.

In generale, queste applicazioni di determinazione della pericolosità dell'individuo sembrano riflettere il dibattito sulla presunta prevedibilità del comportamento criminale, molto vivo negli Stati Uniti ma risolto da circa trent'anni in diversi Paesi europei. Ad esempio, in Italia, l'art. 220, comma 2, cpp esclude espressamente il ricorso alla perizia al fine di stabilire «l'esercizio o la professionalità in un reato, la tendenza a commettere un reato, il carattere e la personalità dell'indagato e, in generale, le qualità psicologiche dell'indagato, indipendentemente dalle cause patologiche». Allo stesso modo, in Francia, la dottrina della "nuova difesa sociale" avanzata da Marc Ancel è stata fondatrice del diritto penale postbellico: un approccio meramente punitivo e deterministico è stato sostituito da un sistema di riabilitazione sociale volto a prevenire la commissione di un reato¹⁴. Un approccio condiviso da una serie di strumenti europei sulle politiche penali che si concentra sugli obiettivi di riabilitazione e reinserimento delle sanzioni penali.

Gli strumenti di giustizia penale dovrebbero pertanto essere concepiti in modo diverso da quello attuale: un approccio "predittivo" fondato sul trattamento statistico di fattori di rischio dovrebbe essere sostituito da un approccio più rispettoso dei diritti umani, in cui il giudice svolge un ruolo fondamentale nell'individualizzazione della sanzione penale. Le tecniche di analisi dei dati potrebbero essere utilizzate dai servizi sociali e da agenti di *probation* per centralizzare e raccogliere informazioni sull'autore di un reato sparse tra diverse istituzioni (informazioni concernenti la formazione, l'occupazione, l'assistenza medico-sociale regolare); queste informazioni potrebbero informare il processo decisionale di un giudice che deve pronunciarsi, a volte, in tempi estremamente brevi (ad esempio, nel contesto delle cosiddette procedure di "comparizione immediata" in Francia o del "processo immediato" in Italia).

13. Si veda, in particolare, www.propublica.org/article/technical-response-to-northpointe.

14. Come nel sistema giuridico italiano, questa filosofia ha portato, ad esempio, alla previsione di misure alternative alla detenzione, come la liberazione condizionale, la libertà vigilata e l'affidamento in prova ai servizi sociali.

3. Verso un indispensabile quadro etico

Come menzionato in introduzione, la Cepej ha appena adottato la prima «Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari», che mette un accento particolare sulle tecniche di trattamento e analisi automatizzate delle decisioni giudiziarie. Si tratta di uno strumento non vincolante, ma che ha il merito di enunciare principi sostanziali e metodologici applicabili sia agli attori privati (come le *start-up* attive sul mercato delle nuove tecnologie applicate ai servizi giuridici – le *legaltech*) sia alle autorità pubbliche. Queste ultime potranno trarre ispirazione dai suddetti principi per definire delle politiche o un quadro normativo in questa materia.

Tra i principi enunciati, il rispetto dei diritti dell'uomo e della non-discriminazione da parte delle applicazioni di intelligenza artificiale riveste un'importanza fondamentale. Si tratta di assicurare, dalla fase della concezione fino all'applicazione pratica, che le soluzioni garantiscano il rispetto dei diritti garantiti dalla Convenzione europea dei diritti dell'uomo e dalla Convenzione n. 108 del Consiglio d'Europa sulla protezione dei dati personali. Il principio di non-discriminazione è espressamente enunciato in ragione della capacità dei trattamenti Intelligenza artificiale – soprattutto nella materia penale – di rivelare discriminazioni esistenti, raggruppando o classificando dati riguardanti persone o gruppi di persone. Gli attori pubblici e privati devono dunque garantire che queste applicazioni non riproducano o aggravino tali discriminazioni e non portino ad analisi o pratiche deterministiche.

Vengono anche prese in considerazione alcune sfide qualitative sul piano della metodologia dell'analisi e del trattamento automatizzato delle decisioni giudiziarie. L'elaborazione dei dati effettuata tramite *machine learning* dovrebbe poter essere effettuata sulla base di originali certificati e l'integrità di tali dati dovrebbe essere garantita in tutte le fasi del trattamento. Inoltre, ci dovrebbe essere una particolare attenzione nella fase della selezione delle decisioni che saranno in seguito trattate con l'apprendimento automatico. Queste dovrebbero essere realmente rappresentative delle diverse realtà su cui il giudice è chiamato a statuire e non corrispondere a griglie di analisi predeterminate (per esempio, i progettisti potrebbero avere la tendenza a scartare decisioni che non si prestino alle correlazioni di sequenze linguistiche da parte dell'apprendimento automatico o decisioni scarsamente motivate). Inoltre, si sottolinea la necessità di disporre di un ambiente tecnologico sicuro per l'archiviazione e l'attuazione di modelli di *machine learning* e di algoritmi.

Di grande importanza è anche il principio di trasparenza delle metodologie e delle tecniche utilizzate nel trattamento delle decisioni giudiziarie. Si pone qui l'accento sulla necessità di rendere accessibili e comprensibili le tecniche di trattamento dei dati, così come di autorizzare delle verifiche (*audits*) esterne da parte di soggetti indipendenti, al fine di individuare eventuali distorsioni. È, altresì, incoraggiato un sistema di certificazione delle diverse applicazioni da parte di queste autorità.

Inoltre, si sottolinea la necessità di non subire passivamente e, al contrario, di rinforzare l'autonomia degli utenti nell'uso di strumenti e servizi di intelligenza artificiale. Il giudice, in particolare, dovrebbe essere in grado di tornare in qualsiasi momento alle decisioni giudiziarie e ai dati che sono stati utilizzati per produrre un risultato, e continuare ad avere la possibilità di discostarsi da esse in considerazione delle specificità del caso concreto. Ogni utente dovrebbe essere informato, in un linguaggio chiaro e comprensibile, della natura vincolante o non vincolante delle soluzioni proposte dagli strumenti di intelligenza artificiale, delle varie opzioni possibili e del suo diritto all'assistenza legale e al ricorso a un tribunale.

Si considera che le sfide poste dall'Intelligenza artificiale nel settore della giustizia potranno essere trattate in maniera più efficace soltanto attraverso una fusione dei saperi: a questo proposito, la creazione di *équipe* pluridisciplinari composte da magistrati, ricercatori in scienze sociali e informatici è fortemente raccomandata, tanto allo stadio di elaborazione che di pilotaggio e di attuazione delle soluzioni proposte. In generale, si raccomanda un approccio graduale nell'integrazione dell'Intelligenza artificiale nei sistemi giudiziari: è essenziale, in effetti, verificare che le applicazioni proposte (spesso da imprese private) abbiano un valore aggiunto per gli utenti e che contribuiscano effettivamente alla realizzazione degli obiettivi del servizio pubblico della giustizia.

L'auspicio della Cepej è che questi principi diventino l'asse portante di una "cybergiustizia" a livello europeo e che possano divenire un riferimento concreto per i magistrati, le autorità politiche o giudiziarie che devono far fronte alla sfida dell'integrazione delle nuove tecnologie fondate sull'Intelligenza artificiale nelle politiche pubbliche o nel lavoro quotidiano. Inoltre, su un piano pratico, questi principi rappresentano una base di comparazione importante per valutare le caratteristiche delle differenti applicazioni di Intelligenza artificiale di cui si valuta l'applicazione nel sistema giudiziario o a livello dei tribunali. La Cepej resta, naturalmente, a disposizione degli Stati membri, dei tribunali e dei rappresentanti delle diverse professioni giuridiche per accompagnarli nell'attuazione dei principi della Carta etica.