

Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale in prospettiva comparata

di Simone Penasa

Abstract: *Artificial intelligence and justice: the fragile balancing between technological reliability and constitutional sustainability in comparative perspective* – The paper highlights the main pros and cons of the use of AI in the judicial context, taking into account different functions that AI can concretely perform. Technological reliability and constitutional sustainability shall represent key-criteria in order to assessing the impact on both fundamental rights and the judiciary function that the use by judges and other public authorities in the judicial environment can provoke. Special attention is dedicated to the existing case-law and legislations from a comparative perspective, with the aim to provide a first classification of different regulatory approaches to the field.

297

Keywords: Artificial intelligence; Predictive justice; Comparative law; Rule of law; Judiciary.

1. Introduzione: per un approccio realista all'utilizzo dell'intelligenza artificiale in ambito giudiziario

Nelle riflessioni che seguiranno, dedicate agli effetti di contesto e di sistema che l'utilizzo in ambito giudiziario di sistemi basati sull'intelligenza artificiale può produrre a livello giuridico, si tenterà di adottare un approccio realista, che si ponga cioè in una linea mediana tra atteggiamenti irenici e reazioni “millenaristiche”. Tale approccio si poggia idealmente su due assunti, che esprimono la natura della relazione tra dimensione giuridica e dimensione tecnologica, da un lato, e l'ordine di precedenza tra queste ultime e, più in generale, tra dimensione assiologica – pur se integrata in senso epistemologico e metodologico dalle acquisizioni tecnico-scientifiche¹ – e dimensione tecnologica, dall'altro lato.

Nel primo senso, si ritiene che l'eventuale sostituzione del giudice con la macchina (giudice robot) nell'esercizio di funzioni giurisdizionali, anche

• L'articolo è stato scritto nell'ambito del Jean Monnet project *Trento AI Laboratory* (TrIAL)

¹ Ho cercato di sviluppare questa prospettiva, da ultimo, in S. Penasa, *Ai confini del diritto pubblico: ibridazione dei metodi e identità epistemologiche nel prisma del primato della Costituzione*, in *Rivista AIC*, 3, 2020, 203-223.

quando tale sviluppo potrà essere tecnicamente realizzabile (general AI²), rappresenti un'ipotesi che l'ordinamento giuridico-costituzionale è in grado di prevedere, gestire ed eventualmente scongiurare (o almeno governare), mantenendo la titolarità del potere di controllo sull'an e il quomodo dell'utilizzo. Nel secondo senso, e al fine di assicurare una disciplina adeguata, efficace ed effettiva, appare necessario ideare e attivare fin d'ora processi decisionali "ibridi"³, in attuazione dell'idea di una evidence-based law⁴, attraverso i quali giungere a una esaustiva conoscenza dell'oggetto di disciplina, delle possibili applicazioni e degli eventuali rischi nella prospettiva della tutela dei diritti e della tenuta dei principi che tradizionalmente orientano l'esercizio della funzione giurisdizionale.

Tale operazione deve avvenire, evidentemente, in modo da assicurare e – dove possibile – promuovere due distinte ma correlate dimensioni relative alla sostenibilità in termini giuridici dell'interazione tra strumenti tecnologici e funzioni pubbliche. Da un lato, si pone l'esigenza di assicurarne la sostenibilità costituzionale, la quale si compone tanto di una dimensione assiologica, relativa all'impatto sui diritti e sulle libertà delle persone coinvolte, quanto di una dimensione propriamente istituzionale, la quale richiede una valutazione degli effetti – fisiologici o potenzialmente distorsivi – dell'impiego di sistemi autonomi sulla natura della funzione giurisdizionale e sul ruolo del giudice all'interno di uno stato costituzionale di diritto⁵. Dall'altro lato, ci si può esprimere in termini di sostenibilità tecnologica degli strumenti normativi concretamente adottati, la quale rappresenta uno sviluppo coerente con la riflessione dottrinale e la giurisprudenza (italiana e internazionale, Corte EDU) in materia di qualità tecnico-scientifica o di

² Su tale concetto, *ex plurimis* R. Susskind, *Online Courts and the Future of Justice*, Oxford, 2019, in particolare 264 ss.

³ La letteratura costituzionalistica sul metodo legislativo da adottare in relazione a leggi science-related è ormai molto sviluppata, si rinvia *ex plurimis* a A. Iannuzzi, *Il diritto capovolto. Regolazione a contenuto tecnico-scientifico e Costituzione*, Napoli, 2018; G. Ragone, *Eine empirische Wende? La Corte costituzionale e le sfide della complessità tecnico-scientifica*, Torino, 2020; D. Servetti, *Riserva di scienza e tutela della salute. L'incidenza delle valutazioni tecnico-scientifiche di ambito sanitario sulle attività legislative e giurisdizionale*, Pisa, 2020; G. D'Amico, *Scienza e diritto nella prospettiva del giudice delle leggi*, Messina, 2008. Per un approccio interdisciplinare, si rinvia alla sezione monografica in materia ospitata dal n. 3 del 2020 di questa *Rivista*.

⁴ Si riferisce a tale concetto C. Casonato, *Evidence Based Law. Spunti di riflessione sul diritto comparato delle scienze della vita*, in *BioLaw Journal-Rivista di BioDiritto*, 1, 2014, 179-208.

⁵ In riferimento alla dottrina costituzionalistica italiana, senza pretesa di esaustività, M. Luciani, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018, 872-893; 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, 87-105; F. Donati, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020, 414-436; C. Colapietro, A. Moretti, *L'Intelligenza Artificiale nel dettato costituzionale: opportunità, incertezze e tutela dei dati personali*, in *BioLaw Journal-Rivista di BioDiritto*, 3, 2020, 359-387; A. D'Aloia, *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLaw Journal-Rivista di BioDiritto*, 1, 2019, 1 ss.; C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *DPCE*, 2019, numero speciale, 101-130.

ragionevolezza scientifica delle leggi “science related”⁶ e richiede – in termini generali⁷ – che sia garantita l’adeguatezza anche dal punto scientifico e tecnologico delle norme previste e la capacità di queste ultime di trovare effettiva attuazione nell’ambito regolato da parte degli operatori concretamente coinvolti⁸.

A tali fini e a fronte di un quadro normativo che, a livello interno e comparato, sta esprimendo le prime iniziative legislative finalizzate a regolare l’utilizzo di tali tecnologie nell’ambito dei tradizionali poteri pubblici, e che, a livello sovranazionale e internazionale, esprime una per certi versi sorprendente tendenza alla produzione di strumenti tanto di soft law quanto di hard law⁹, è parso opportuno fondare l’analisi che seguirà su di un approccio induttivo, che prenda avvio dall’analisi di alcuni casi giurisprudenziali particolarmente rilevanti. Prima di avviare tale percorso, appare necessario fornire alcune coordinate generali utili a orientare l’analisi critica delle esperienze comparate che saranno analizzate nel presente scritto.

2. IA e giustizia: l’esigenza di una mappa concettuale, assiologica e funzionale per orientare il giudizio di compatibilità

Dovendo fornire un quadro complessivo, pur se sintetico, della relazione tra intelligenza artificiale e giustizia, risultano molteplici le prospettive dalle quali tale rapporto può essere osservato e valutato. Dal punto di vista della natura di tale relazione, in particolare del tipo di effetti che l’impiego di tali tecnologie può produrre sulla natura e sulle modalità di esercizio di funzioni

⁶ Su tale concetto, A. Iannuzzi, *Leggi “science driven” e CoViD-19. Il rapporto fra politica e scienza nello stato di emergenza sanitaria*, in *BioLaw Journal-Rivista di BioDiritto*, 1, 2020, Special Issue, 128-134. Sul principio di ragionevolezza scientifica delle leggi, oltre alle opere citate alla nota 4, *ex plurimis* C. Di Costanzo, A. Simoncini, *Il contributo della Corte costituzionale allo sviluppo del biodiritto*, in *BioLaw Journal-Rivista di BioDiritto*, 2, 2019, Special Issue, 205-240; L. Busatta, *Tra scienza e norma: il fattore scientifico come oggetto, strumento e soggetto della regolazione*, in *Costituzionalismo.it*, 1, 2021, in particolare 154 ss.; D. Zanoni, *Razionalità scientifica e ragionevolezza giuridica a confronto in materia di trattamenti sanitari obbligatori*, in *Costituzionalismo.it*, 2, 2020; S. Troilo, *La Corte costituzionale e le questioni tecnico-scientifiche in materia sanitaria*, in *Quaderni costituzionali*, 2, 2020, 355-369; U. Adamo, *Materia “non democratica” e ragionevolezza della legge*, in *Consulta Online*, 1, 2018, 300 ss.; E. Cheli, *Scienza, tecnica e diritto: dal modello costituzionale agli indirizzi della giurisprudenza costituzionale*, in *Rivista AIC*, 1, 2017; C. Casonato, *La scienza come parametro interposto di costituzionalità*, in *Rivista AIC*, 2, 2016, 1-11; A. Morrone, *Ubi scientia ibi iura. A prima lettura sull’eterologa*, in *Forum di Quaderni costituzionali*, 11 giugno 2014; S. Penasa, *La “ragionevolezza scientifica” delle leggi nella giurisprudenza costituzionale*, in *Quaderni costituzionali*, 4, 2009, 817-841.

⁷ Si rinvia al paragrafo conclusivo per considerazioni più approfondite.

⁸ Sul punto, le puntuali osservazioni di A. Simoncini, *Sistema delle fonti e nuove tecnologie. Le ragioni di una ricerca di diritto costituzionale, tra forma di stato e forma di governo*, in *Osservatorio sulle fonti*, 2, 2021, 730.

⁹ Per un’approfondita analisi di tale dimensione della governance dell’IA, L. Parona, *Prospettive europee e internazionali di regolazione dell’intelligenza artificiale tra principi etici, “soft law” e “self regulation”*, in *Rivista della regolazione dei mercati*, 1, 2020, 70-93.

giurisdizionali, nonché sui soggetti titolari delle medesime, è stata proposta una tassonomia degli effetti che l'utilizzo della tecnologia può produrre su un sistema giudiziario, secondo un ordine crescente di incidenza e riconfigurazione della natura della funzione giudiziaria e del giudice: una tecnologia può assistere, cioè “to inform, support and advise” (“supportive technology”); a un livello intermedio di complessità tecnica e criticità almeno potenziale in termini giuridici, si individuano le “replacement technologies”, le quali sono in grado di “replace functions and activities that were previously carried out by humans”; infine, si può considerare ciò che viene definito “disruptive technology”, categoria nella quale rientrano quei sistemi autonomi in grado di “change the way that judges work and provide for very different forms of justice (...), particularly where processes change significantly and predictive analytics may reshape the adjudicative role”¹⁰.

La tassonomia appena richiamata si fonda sulla distinzione non tanto dei diversi compiti, o della diversa natura, che questo tipo di sistemi autonomi possono svolgere o assumere, quanto piuttosto sul possibile impatto che questi ultimi sono in grado di produrre sulla concezione della funzione giudiziaria e sull'immagine stessa che il giudice esprime nell'esercitarla nei confronti del contesto sociale di riferimento (anche solo in termini di apparenza). In tale prospettiva, è destinato ad assumere una particolare centralità, accanto al tipo di funzione concretamente attribuita a un determinato sistema “intelligente”, il livello di autonomia espresso dalla tecnologia adottata. Il supporto all'interno dell'attività giudiziaria, può avvenire infatti attraverso sistemi caratterizzati da un livello crescente di autonomia, dai più tradizionali e consolidati strumenti di “decision-support”, a sistemi esperti che richiedono comunque una interazione con il titolare della funzione (“human-in-the-loop”), fino a sistemi in grado di sostituire completamente il titolare nello svolgimento di attività e funzioni di natura giudiziaria, provocando la “scomparsa” dell'agente umano dal processo decisionale¹¹.

L'aumento del livello di autonomia dei sistemi esperti finisce pertanto con il produrre due effetti che devono essere affrontati da punto di vista normativo: la possibilità tecnica di delegare a tali sistemi spazi decisionali più o meno ampi, tradizionalmente riservati ai titolari “umani” delle funzioni corrispondenti, da un lato¹²; la sostenibilità costituzionale di tale eventuale

¹⁰ T. Sourdin, *Judge v Robot? Artificial Intelligence and Judicial Decision-Making*, in *UNSW Law Journal*, 4, 2018, 1117 ss.; cfr. anche A. D. Reiling, *Courts and Artificial Intelligence*, in *International Journal for Court Administration*, 2, 2020, 3-6, nel quale si distingue tra le funzioni di “organising informations”, “advise” e “predictions”.

¹¹ Sulla sostenibilità costituzionale di una totale sostituzione del giudice da parte di sistemi esperti, *ex plurimis* e con specifico riguardo ai principi costituzionali in materia di magistratura, F. Donati, *Intelligenza artificiale e giustizia*, cit., 429 ss., il quale richiama opportunamente la giurisprudenza costituzionale italiana in materia di automatismi legislativi (su tale aspetto, anche A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal-Rivista di BioDiritto*, 1, 2019, 81 ss.).

¹² M. Luciani, *La decisione giudiziaria robotica*, cit., 872 s., individua tre fondamentali

scelta, che non potrà che basarsi – almeno relativamente all’an¹³ – su una esplicita manifestazione di volontà politica del legislatore.

A fronte di tale quadro complesso, nel quale si esprimono valutazioni relative alla fattibilità tecnica, legittimità costituzionale e opportunità politica dell’utilizzo di tali sistemi in ambito giudiziario, in dottrina è stata proposta una categorizzazione che si fonda su tre tipi di criticità potenzialmente derivanti da tale interazione¹⁴, l’analisi delle quali porta l’Autore a concludere “per l’esclusione di una sostituzione del giudice e per l’opportunità, invece, di affidare alla AI un ruolo di “supporto necessariamente ponderato” in riferimento all’esercizio della funzione giurisdizionale (...)”¹⁵. Se, interpretando tale approccio, le criticità di natura tecnologica (errori e bias, neutralità e trasparenza¹⁶) e giuridica (diritto di difesa e contraddittorio, appello, stasi giurisprudenziale) potrebbero definirsi come ‘contestuali’, in quanto potenzialmente superabili attraverso l’evoluzione delle acquisizioni e delle applicazioni in ambito di IA – da un lato – e l’adeguamento rispetto all’utilizzo nel processo di tali tecnologie delle tradizionali regole sostanziali e processuali¹⁷, residuerebbero in ogni caso criticità ‘strutturali’ – di ordine ontologico, in quanto direttamente derivanti dall’esigenza di assicurare la natura della funzione giurisdizionale e del ruolo del giudice riconducibili ai principi del costituzionalismo – che non sarebbero superabili attraverso innovazioni tecnologiche o iniziative di natura giuridica¹⁸. In tale ultima prospettiva, infatti, “rimarrebbe sempre la distanza fra il genere umano e quello artificiale”¹⁹.

A fronte di un quadro che si presenta tecnologicamente e giuridicamente complesso, in quanto l’intreccio tra IA e giustizia coinvolge e affonda nelle radici delle garanzie costituzionali tipicamente riconosciute in tale ambito, e a una riflessione dottrinale già molto sviluppata sugli aspetti teorico-generalisti, appare opportuno dedicare la seconda parte di questo contributo all’analisi delle prime iniziative normative finalizzate a governare

ragioni – liberare dal peso del lavoro, prestazioni più efficienti, garantire certezza giuridica – e distingue tra sostituzione integrale e parziale della macchina all’essere umano (Ivi, 873).

¹³ Cfr. A. Simoncini, *Sistema delle fonti e nuove tecnologie. Le ragioni di una ricerca di diritto costituzionale, tra forma di stato e forma di governo*, cit., 727-731; e A. Celotto, *Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto*, in *Analisi Giuridica dell’Economia*, 1, 2019, 47-59.

¹⁴ R. Susskind, *Online Courts and the Future of Justice*, cit., 278-279, individua cinque livelli di rilevanza riferibili alla questione della sostituibilità del giudice umano da parte della macchina: se ciò sia “technically possible”, “morally acceptable”, “commercially viable”, “culturally sustainable” e “jurisprudentially coherent”.

¹⁵ C. Casonato, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in DPCE Online, 3, 2020, 3371.

¹⁶ Ivi, 3376.

¹⁷ Ivi, 3384.

¹⁸ Ibidem, nel quale si afferma che “talune qualità umane ritenute necessarie per il giudice non potranno mai essere sostituite da componenti artificiali per quanto avanzati e sofisticati”.

¹⁹ Ivi, 3385.

l'utilizzo dell'IA a fini di giustizia. Anche alla luce della natura ancora “embrionale” e per certi tratti sperimentale²⁰ delle riforme legislative adottate allo stato, si aprirà l'analisi alla dimensione giurisprudenziale.

3. “La” o “le” discipline dell'IA in ambito giudiziario? *Risk assessment tools*, “predictive” justice, profiling dei giudici, supportive AI

Nell'adottare un approccio casistico, si è proceduto alla selezione di ambiti applicativi particolarmente paradigmatici, nei quali le potenzialità e i rischi derivanti dall'utilizzo di sistemi esperti emergono in modo particolarmente evidente, da un lato, e sono stati allo stesso tempo oggetto di disciplina legislativa o di scrutinio giurisdizionale, dall'altro lato. Richiamando le tassonomie relative alle possibili funzioni esercitabili da tali tecnologie che sono state richiamate nel secondo paragrafo, i casi analizzati sono suscumbibili a tipi di “supportive” e “disruptive technologies”, senza giungere a interessare esempi di completa sostituzione del giudice umano da parte della tecnologia²¹.

L'analisi comparata non può che prendere avvio dal caso COMPAS, che indubitabilmente deve essere considerato come il leading case in materia²². In questa sede ci si vuole concentrare sulla valutazione relativa alla legittimità dell'utilizzo di sistemi di valutazione del rischio di recidiva e di pericolosità sociale svolta dalla Corte Suprema del Wisconsin²³ nel caso Loomis. La Corte Suprema, come noto, affermando la necessità di assicurare anche in tale ambito l'effettività delle garanzie riconducibili alla “due process clause”, ne specifica le finalità ammissibili e gli effetti sui poteri del giudice, individuandone di conseguenza le condizioni di utilizzo.

In relazione al primo aspetto – le finalità per le quali tali risk assessment tools possono essere utilizzati²⁴ – la Corte sottolinea che le valutazioni (“risk scores”) relative alla pericolosità sociale e al rischio di

²⁰ Sul concetto di experimental laws, S. Ranchordas, *Experimental Regulations and Regulatory Sandboxes: Law without Order?*, in *University of Groningen Faculty of Law Research Paper*, 10, 2021, 30 settembre 2021.

²¹ R. W. Campbell, *Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning*, in *Colorado Technology Law Journal*, 2, 2020, 543 ss., individua due modelli generali attraverso i quali è possibile “bringing AI directly into the judicial decision process without *ex ante* human guidance”.

²² Tra i numerosi commenti, J. Ulenaers, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, in *Asian Journal of Law and Economy*, 11, 2, 2020, 7 ss.; A. Simoncini, *L'algoritmo incostituzionale*, cit., 71-73; G. De Minico, *Towards an “Algorithm Constitutional by Design”*, in *BioLaw Journal-Rivista di BioDiritto*, 1, 2021, 399 ss.

²³ Supreme Court of Wisconsin, *State of Wisconsin v. Eric L. Loomis*, caso n. 2015AP157-CR, 5 aprile–13 luglio 2016.

²⁴ Per un'analisi del sistema COMPAS e del suo equivalente britannico HART, M. Oswald et al., *Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality*, in *Information & Communications Technology Law*, 2, 2018, 223-250.

recidiva prodotte da COMPAS non possono essere utilizzate dal giudice al fine di determinare *l'an* e il *quantum* della pena detentiva. Dopo averne limitato in modo almeno teoricamente severo l'ambito di utilizzo alla determinazione delle modalità di esecuzione della pena più adeguate a ciascun individuo²⁵, la Corte Suprema riconosce le potenzialità di tali strumenti al fine di migliorare la valutazione del giudice, la ponderazione e l'applicazione delle altre prove nella formulazione di una decisione individualizzata che risulti appropriata ad ogni imputato. Al contempo, ne delimita l'impatto, chiarendo che "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"²⁶. In tal senso, la Corte Suprema richiama la necessità che il giudice eserciti in modo adeguato la propria discrezionalità nel momento in cui si trovi ad applicare il livello di rischio determinato da COMPAS in relazione a ciascun imputato²⁷, tenuto conto anche del fatto che l'algoritmo è in grado di identificare il livello di presunta pericolosità sociale di gruppi di popolazione, ma non quello specifico di una determinata persona²⁸.

Occorre pertanto evitare ogni automatica ricezione, sussistendo un dovere (professionale) del giudice di spiegare (motivare) i fattori sui quali si sia basato per determinare la condanna, indipendentemente dall'impiego dei dati ricavati da COMPAS, in quanto secondo la Corte Suprema "A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing"²⁹. La Corte Suprema, basandosi anche sulle caratteristiche tecniche del sistema utilizzato, afferma che – se svolto in modo appropriato ("properly") – l'utilizzo di strumenti simili a COMPAS non si pone in contrasto con il diritto dell'imputato al giusto processo ("due process") e individua una serie di condizioni e cautele in grado di orientare il giudice verso un uso appropriato dei medesimi, ad esempio l'esistenza di una informazione adeguata relativa alle caratteristiche specifiche e ai potenziali rischi di discriminazione e stigmatizzazione derivanti dall'utilizzo di tali sistemi.

A tal riguardo, si richiama un caso deciso dalla Corte Suprema canadese³⁰, relativo all'utilizzo da parte del Correctional Service of Canada (CSC) di strumenti di valutazione del rischio di recidiva e di psicopatologia nei confronti di un detenuto appartenente alla minoranza indigena storicamente

²⁵ Anche in considerazione della pluralità di funzioni che tradizionalmente vengono attribuite a una sanzione penale (§ 96-97).

²⁶ Ivi, § 92.

²⁷ Ivi, § 71.

²⁸ Ivi, § 74.

²⁹ Ivi, § 99.

³⁰ Supreme Court, *Ewert v Canada*, 2018 SCC 30. T. Scassa, *Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making*, in *University of British Columbia Law Review*, 54, 1, 2021, 251-298, la quale si esprime in termini di "obligation to question the reliability of data that informs decisions".

residente sul territorio canadese. Alla luce dell'esistenza di possibili bias cognitivi interni al sistema di risk assessment in grado di produrre effetti almeno potenzialmente discriminatori³¹, la Corte Suprema sancisce il mancato rispetto degli standard previsti a livello legislativo al fine di assicurare l'accuratezza e la completezza – in termini più complessivi, l'affidabilità – dei dati utilizzati, ritenendo che la clausola generale che impegna le autorità competenti a svolgere “all reasonable steps” per assicurarne l'effettivo rispetto debba essere interpretata alla luce del contesto al quale risulta applicabile.

La Corte Suprema concretizza tale standard generale individuando un dovere della pubblica amministrazione di svolgere indagini e ricerche finalizzate ad accertare se – e in quale misura – tali sistemi siano soggetti a “cross-cultural variance” quando vengano applicati a detenuti appartenenti a una minoranza etnica. Lo svolgimento di tale attività di indagine rappresenta quindi una condizione minima di legittimità dell'utilizzo di tali strumenti, dalla quale deriva un contestuale dovere di modificarne l'utilizzo in modo da eliminare possibili rischi di pregiudizio per la popolazione carceraria indigena derivante³². Lo scrutinio si orienta verso la verifica del rispetto – da parte dei titolari delle funzioni pubbliche – di requisiti di natura procedurale, risultando la legittimità dell'utilizzo condizionata a una preventiva verifica della affidabilità degli algoritmi adottati, la validità dei quali deve essere valutata anche in termini di possibili effetti discriminatori³³.

Tale tendenza trova conferma nella sentenza della Court of Appeal britannica, relativa all'utilizzo da parte della polizia di sistemi di riconoscimento facciale, basati su algoritmi, a fini di prevenzione e repressione dei crimini³⁴. Anche la Court of Appeal, infatti, afferma il dovere per le autorità competenti di valutare *ex ante* “the potential impact of a new policy which may appear (...) to be neutral but which may turn out in fact to have a disproportionate impact on certain sections of the population”³⁵. La Court sembra compiere un passo ulteriore, chiarendo come tale valutazione

³¹ I sistemi di previsione del rischio oggetto del ricorso sono testati e sviluppati su una popolazione carceraria prevalentemente non appartenente alla minoranza indigena e nessuno studio è stato svolto al fine di confermarne la validità anche in caso di applicazione a detenuti appartenenti a tale gruppo sociale.

³² Ivi, § 67.

³³ G. De Minico, *Towards an “Algorithm Constitutional by Design”*, cit., 402, rileva che “The Canadian Supreme Court, ruling in Ewert’s favour, determined that, without evidence of the algorithm being free of cultural bias, it was unjust to use this tool on indigenous inmates”.

³⁴ Court of Appeal, R (Bridges) v. The Chief Constable of South Wales Police et al., [2020] EWCA Civ 1058, sulla quale J. Purshouse, L. Campbell, *Automated facial recognition and policing: A Bridge too far?*, in *Legal Studies*, online, 2021, 1-19. Per un'analisi della decisione della High Court che ha preceduto quella della Court of Appeal, A. Pin, *Non esiste la “pallottola d'argento”: l'Artificial Face Recognition al vaglio giudiziario per la prima volta*, in *DPCE Online*, 4, 2019, 3075-3081.

³⁵ Court of Appeal, R (Bridges) v. The Chief Constable of South Wales Police et al., cit., § 179.

non possa essere surrogata da garanzie ulteriori, quali la verifica *ex post* di un operatore umano (principio dello “human in the loop”) o la presenza nel sistema di correttivi statistici che tengano in considerazione l’elemento del genere e dell’origine etnica.

A livello legislativo, sono due gli ambiti rispetto ai quali si registrano le iniziative più innovative: la disciplina di “pretrial risk assessment tools”, da un lato, e di sistemi di judge’s profiling finalizzati a prevedere in termini probabilistici l’esito del giudizio, dall’altro lato.

Rispetto al primo ambito, si richiama la riforma avvenuta nel 2019 nello Stato dell’Idaho, negli Stati Uniti, che rappresenta il primo esempio di disciplina legislativa in materia a livello comparato³⁶. La legge rappresenta il tentativo di garantire il rispetto dei principi di trasparenza, pubblicità, non discriminazione e del diritto a un processo equo nel caso in cui si utilizzino strumenti di valutazione del rischio in una fase precedente al processo. In termini di trasparenza, si prevede che ogni documento, dato o documentazione che siano stati utilizzati dal produttore per costruire o validare questo tipo di sistemi, così come ogni altra fonte che ne descriva l’utilizzo o la validazione sia reso pubblico, al fine di consentire attività di ispezione, revisione e controllo (“inspection, auditing and testing”). Relativamente all’accessibilità delle informazioni, si vieta ai produttori o a chi se ne avvalga di addurre motivi di privativa industriale o di proprietà intellettuale al fine di impedire la conoscenza di tali dati e informazioni, in relazione a casi di natura sia penale che civile. Nella prospettiva della tutela delle parti processuali, si afferma il diritto dell’imputato di riesaminare (“review”) tutti i calcoli svolti e i dati utilizzati al fine di misurare il livello di rischio dello stesso, anche quando questo tipo di valutazione sia stata svolta non direttamente dalla corte ma da un perito.

In relazione al profiling dei giudici, la Francia rappresenta il primo ordinamento giuridico nel quale sia stato previsto un divieto assoluto di svolgere questo tipo di attività, prevedendo quale sanzione una pena detentiva fino a cinque anni. Nello specifico, l’art. 33 della legge n. 2019-222³⁷ ha previsto che i dati d’identità dei giudici non possono essere riutilizzati con lo scopo o l’effetto di valutare, analizzare, comparare o prevedere le loro pratiche professionali reali o presunte. Si tratta, di tutta evidenza, di una opzione legislativa che adotta un approccio assolutamente proibizionista, che – per le sue potenziali ricadute negative sul principio di uguaglianza davanti alla legge e sul diritto a un processo equo – è stata oggetto di scrutinio da parte del Conseil Constitutionnel francese³⁸. Quest’ultimo ha giudicato legittima la disposizione, in quanto il legislatore francese attraverso la predisposizione di

³⁶ Idaho Criminal Code sec. 19-1910, 2019. Per un’analisi del caso, B. Winters, *Openish Government Laws and the Disparate Racial Impact of Criminal Justice Technologies*, in *The Law Review of the Franklin Pierce Center for Intellectual Property*, 3, 2021, 537 ss.

³⁷ Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice.

³⁸ Conseil Constitutionnel, Décision n° 2019-778 DC, 21 marzo 2019.

un divieto assoluto ha voluto assicurare il corretto funzionamento del servizio giustizia, il quale avrebbe potuto subire indebite alterazioni, causate da pressioni o strategie finalizzate alla scelta di una specifica giurisdizione, effettuata sulla base della profilazione dei giudici³⁹. Inoltre, il Conseil non ha rilevato alcuna eventuale differenziazione ingiustificata, né alcuna violazione del diritto a un processo equo, in quanto la disciplina garantisce un adeguato bilanciamento tra i diritti delle parti processuali⁴⁰. In termini generali, i giudici costituzionali francesi hanno fatto prevalere esigenze di natura ordinamentale, sacrificando l'operatività di uno strumento potenzialmente disponibile per le parti processuali a ragioni di equità nell'accesso al servizio giustizia e – seppur indirettamente – di garanzia dei giudici rispetto anche a solo potenziali pressioni o condizionamenti in grado di incidere sulle concrete modalità di esercizio della funzione giurisdizionale⁴¹.

Il caso del divieto francese all'attività di profiling dei giudici consente di sottolineare un ulteriore variabile decisiva nell'orientare il giudizio relativo all'ammissibilità di sistemi autonomi basati su IA. In questo caso, l'algoritmo che produce la profilazione rappresenta un servizio disponibile per le parti processuali, integrando quindi la strategia processuale di parte, e proprio sulla base della funzione concretamente svolta, il legislatore vi associa un disvalore sanzionabile penalmente, in quanto potenzialmente in grado di alterare l'equilibrio interno alla funzione giurisdizionale. Ci si potrebbe chiedere se, al mutare della funzione attribuita, la valutazione debba restare la medesima, o possa mutare, ad esempio nel caso in cui il medesimo strumento algoritmico venga utilizzato quale risorsa finalizzata alla formazione e all'aggiornamento dei giudici, svolgendo quindi una funzione interna alla magistratura. Si tratta di un aspetto fondamentale, che dimostra paradigmaticamente come il giudizio di ammissibilità rispetto a tali strumenti non possa essere fornito in termini assoluti, ma debba essere contestualizzato, in quanto necessariamente riferito alla e calibrato sulla funzione concretamente svolta da questi ultimi, l'individuazione della quale non può che essere rimessa alla discrezionalità e alla responsabilità del Parlamento. È la funzione attribuita, quindi, non la tecnologia adottata che diviene giuridicamente rilevante e che orienta il giudizio di ammissibilità della seconda rispetto alla prima.

4. Intelligenza artificiale e giustizia: integrazione (della prima) o sostituzione (della seconda)?

Dai casi analizzati emergono due esigenze concomitanti, la cui rilevanza dal punto di vista giuridico dipende dalle funzioni concretamente attribuite ai

³⁹ Ivi, § 93.

⁴⁰ Ivi, § 94.

⁴¹ B. Galgani, *Considerazioni sui "precedenti" dell'imputato e del giudice al cospetto dell'IA nel processo penale*, in *Sistema penale*, 4, 2020, in particolare 91 ss., la quale inquadra la questione anche nella prospettiva della giurisprudenza della Corte EDU in materia di art. 6 CEDU (diritto a un tribunale indipendente e imparziale).

sistemi intelligenti in ambito giudiziario. Da un lato, è necessario assicurare la tutela dei diritti fondamentali delle persone coinvolte, tanto in sede processuale (caso Loomis), quanto rispetto alla funzione di prevenzione dei reati (utilizzo di sistemi di riconoscimento facciale) e di esecuzione della pena (caso Ewert della Corte Suprema canadese). Dall'altro lato, occorre garantire il rispetto dei principi costituzionali relativi all'esercizio della funzione giurisdizionale e dello status del giudice (il divieto di profilazione dei giudici previsto dall'ordinamento francese), in particolare in termini di indipendenza e imparzialità⁴².

L'elemento che accomuna i casi considerati, sia dal punto legislativo che giurisprudenziale, è rappresentato dall'esigenza – che, se non soddisfatta a livello legislativo, tende a imporsi in sede rimediabile – di assicurare, l'affidabilità dei sistemi adottati, la quale deve essere riscontrata in termini di sicurezza e appropriatezza non soltanto dal punto di vista tecnologico, ma anche assiologico. L'affidabilità tecnica deve integrarsi con la sostenibilità assiologica, nel senso che tali strumenti non devono meramente favorire una maggiore efficienza e sicurezza nello svolgimento delle funzioni alle quali risultano associati, ma devono anche assicurare l'effettività e la certezza del rispetto dei diritti fondamentali coinvolti, i quali non possono risultare lesi – anche solo in termini potenziali, di rischio – come conseguenza diretta, anche se involontaria, dell'impiego dei medesimi.

Una affidabilità tecnologica assiologicamente orientata, quindi, che si pone quale condizione di legittimità dell'utilizzo di tali sistemi, e che può concretizzarsi in un dovere per l'autorità competente di svolgere una valutazione *ex ante* dell'impatto prevedibile sui diritti coinvolti, anche attraverso una verifica indipendente o lo svolgimento di studi e analisi finalizzate a comprendere se e in quale misura i sistemi impiegati siano soggetti a errori o bias culturali o di altra natura. In tale prospettiva, l'intervento legislativo appare necessario, quantomeno al fine di selezionare e tipizzare le funzioni che possono essere svolte attraverso sistemi automatici e individuare i principi generali e le garanzie che devono essere rispettati non solo in sede applicativa ma fin dalla fase di ideazione, progettazione e produzione dei medesimi⁴³.

Entro tale quadro, sono ipotizzabili diversi approcci normativi⁴⁴. Un primo modello esprime una tendenza proibizionista, la quale risulta finalizzata a impedire in via assoluta l'utilizzo di determinati sistemi

⁴² Cfr. le argomentazioni del Conseil Constitutionnel francese nel caso richiamato nel paragrafo precedente.

⁴³ A. Simoncini, *L'algoritmo incostituzionale*, cit., 87-88, si riferisce al concetto di "diritto costituzionale ibrido", il quale esprime l'esigenza di "recepire i valori del diritto costituzionale sin nella fase di progettazione delle macchine", essendo a tal fine necessario un diritto costituzionale "che conosca e sappia esprimersi nel linguaggio della tecnologia (...)".

⁴⁴ Per una panoramica, pur se limitata al contesto statunitense, B. Winters, *Open-ish Government Laws and the Disparate Racial Impact of Criminal Justice Technologies*, in *The Law Review of the Franklin Pierce Center for Intellectual Property*, 3, 2021, 535 ss.

autonomi. La disciplina francese in materia di profilazione dei giudici è suscumbibile in tale approccio, così come – in riferimento al medesimo ordinamento – l’analoga disposizione secondo cui “nessuna decisione giudiziaria che comporti una valutazione del comportamento di una persona può essere basata su un trattamento automatizzato di dati personali destinati a valutare alcuni aspetti della personalità di tale persona”⁴⁵.

Un modello alternativo ha come obiettivo non di impedirne, ma di disciplinarne l’utilizzo, il quale risulta condizionato al rispetto di requisiti, criteri e limiti predeterminati. In tal senso, è possibile riferirsi alla Directive on Automated Systems emanata dal governo canadese⁴⁶, la quale prevede – tra l’altro – la possibilità per la pubblica amministrazione di avere accesso e di testare il sistema utilizzato, nel caso ciò risulti necessario anche nell’ambito di un processo, prevedendo la possibilità di autorizzare anche soggetti terzi a rivedere e verificare tali elementi (§ 6.2.5.2 e 3); e il dovere di svolgere un “Algorithmic Impact Assessment”, prima dell’inizio della produzione, anche al fine di verificare l’assenza di bias involontari riconducibili ai dati utilizzati (§ 6.3.1).

Una soluzione intermedia – ibrida – può consistere in un modello che incorpori elementi riconducibili ai due approcci appena richiamati. Questo sembra essere il caso della recente Proposta di Regolamento a livello di Unione Europea⁴⁷, nella quale, attraverso l’adozione del criterio del livello di rischio associabile ai diversi tipi di sistemi autonomi, si prevede un divieto assoluto di utilizzo esclusivamente ad alcuni sistemi ai quali viene associato un livello di “rischio inaccettabile” e solo nei casi in cui tale rischio possa provocare danni fisici o psicologici⁴⁸, mentre per la commerciabilità di sistemi ad “alto rischio” si prevede il rispetto di requisiti predeterminati dalla proposta, da verificare attraverso un procedimento di verifica di conformità⁴⁹.

Ad eccezione dell’approccio totalmente proibizionista, la tendenza degli interventi normativi ai quali si è fatto riferimento sembra orientata verso una disciplina di natura “procedurale”. Ciò caratterizza anche la giurisprudenza analizzata, dalla quale sembra derivare un generale dovere – tanto in sede attuativa, quanto regolativa – di adottare specifiche misure finalizzate a monitorare, identificare ed eventualmente eliminare (o almeno ridurre) eventuali errori o bias che possano produrre effetti distorsivi o lesivi

⁴⁵ Loi n. 78-17 du 6 janvier 1978, art. 10.1, introdotto dalla Loi n- 2018-493 du 20 juin 2018.

⁴⁶ C. Castets-Renard, *Human Rights and Algorithmic Impact Assessment for Predictive Policing*, in H.-W. Micklitz et al. (a cura di), *Constitutional Challenges in the Algorithmic Society*, Cambridge, 2022, 106 ss.

⁴⁷ Per un primo commento sistematico da parte della dottrina italiana, C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di Regolamento dell’Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal-Rivista di Biodiritto*, 3, 2021, 415-437.

⁴⁸ Ivi, 423.

⁴⁹ Ivi, 425 ss. Cfr. anche California, lo Automated Decision Systems Accountability Act (AB 13).

di diritti fondamentali e principi costituzionali, intervenendo sui patterns e sui dati utilizzati da tali sistemi – ove possibile – o interrompendone, in fase attuativa, o vietandone, a livello legislativo, l'impiego.

Anche alla luce della distinzione tra diverse funzioni⁵⁰ e criticità che possono caratterizzare l'utilizzo di AI-based systems in ambito giudiziario, nel momento in cui risulti possibile eliminare – a livello tecnologico – l'eventualità di errori strutturali o bias cognitivi⁵¹ e sia garantito – dal punto di vista della tutela dei diritti individuali che vengono in rilievo⁵² – un ragionevole bilanciamento con gli interessi pubblici perseguiti attraverso l'utilizzo di sistemi autonomi, un eventuale utilizzo in funzione di supporto⁵³ e integrazione delle tradizionali funzioni di pubblica sicurezza, di prevenzione e repressione dei crimini e di concreto esercizio della funzione giurisdizionale potrebbe non comportare criticità di natura ontologica, dal momento che non sembrano in grado di incidere – se adeguatamente governati – sulla natura della funzione né sull'assetto costituzionale che tipicamente la caratterizza⁵⁴. Come affermato dalla Corte Suprema del Wisconsin nel caso Loomis, “if used properly” e nel rispetto di un quadro normativo che preveda limiti e garanzie adeguate tanto da un punto di vista tecnologico quanto costituzionale, questi strumenti possono contribuire a un più efficace, efficiente ed equo⁵⁵ esercizio di funzioni pubbliche.

In ogni caso, un intervento legislativo rappresenta comunque una condizione necessaria di legittimità, dal momento che – anche nel caso in cui l'evoluzione tecnologica consenta di superare le criticità relative a errori e bias – in assenza di una adeguata base legislativa le autorità giudiziarie godrebbero di una eccessiva discrezionalità rispetto alla determinazione

⁵⁰ Interessanti in tal senso le riflessioni svolte da J. Morison, A. Harksens, *Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making*, in *Legal Studies*, 39, 2019, 618-635, e in particolare il concetto di “semi-automated justice”.

⁵¹ Sul concetto di “digital discrimination”, anche nella prospettiva degli strumenti utilizzabili al fine di affrontarne gli effetti sui diritti delle persone, N. Criado, J. M. Such, *Digital Discrimination*, in K. Yeung, M. Lodge (eds.), *Algorithmic Regulation*, Oxford, 2019, 82-97.

⁵² Sui diritti umani di nuova generazione, “appartenenti all'essere umano nei confronti delle applicazioni della AI”, C. Casonato, *Intelligenza artificiale e giustizia: potenzialità e rischi*, cit., 3385 ss.

⁵³ R. W. Campbell, *Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning*, cit., 342.

⁵⁴ Cfr. però A. Završnik, *Algorithmic justice: Algorithms and big data in criminal justice settings*, in *European Journal of Criminology*, 5, 2021, 633, il quale si interroga in modo critico rispetto alla opportunità di svolgere operazioni di “de-biasing”, in quanto “de-biasing entails that inherently political decisions are to be made, for example as to what is merely gendered and what is sexist language that needs to be ‘cleaned’, or what is hate speech targeting minorities and which differential treatment should be deemed to be discriminatory”.

⁵⁵ «A properly built algorithm can reduce crime and jail populations while simultaneously reducing racial disparities», secondo J. Kleinberg et al., *Human Decisions and Machine Predictions*, in *The Quarterly Journal of Economics*, 1, 2018, 241.

delle concrete caratteristiche di tale utilizzo⁵⁶.

Il panorama sembra invece destinato a cambiare nel caso in cui le funzioni svolte da tali sistemi pongano questioni di natura ontologica, le quali non risultino sanabili né a livello tecnologico né a livello normativo, in quanto in grado di incidere anche solo potenzialmente sullo statuto costituzionale delle funzioni giudiziarie e dei giudici. In questi casi, nei quali è possibile sussumere – evidentemente con livelli di criticità differenziati – la profilazione dei giudici e la totale sostituzione di questi ultimi da parte di sistemi autonomi (Francia), l'impiego di sistemi autonomi appare difficilmente compatibile con la particolare caratterizzazione costituzionale della funzione giudiziaria, soprattutto se intesa alla luce della garanzia da indebite pressioni esterne, del principio secondo cui i magistrati si distinguono solo per diversità di funzioni e dell'imparzialità e della soggezione soltanto alla legge del giudice, che devono essere assicurati anche in termini di apparenza.

Simone Penasa
Dip.to di Giurisprudenza
Università degli Studi di Trento
simone.penasa@unitn.it

⁵⁶ In tal senso, J. Purshouse, L. Campbell, *Automated facial recognition and policing: A Bridge too far?*, cit., 19, ritengono che «(...) law should stipulate clearly whether or not it is permissible for police to deploy the technology and provide guidance on the circumstances in which it is appropriate to use the technology before it is operationalized».