OPACO, TROPPO OPACO, QUASI OSCURO.
L'INTELLIGENZA ARTIFICIALE ALLA PROVA DEL REGOLAMENTO UE 2024/1689*

Claudio de Giacomo**

SOMMARIO: 1.- Premessa; 2.- L'analisi predittiva delle sentenze sulla base dei dati; 3.- Il caso Loomis e il sistema Compas; 4.- Gli sviluppi del dibattito; 5.- Il Regolamento UE 2024/1689 tra indirizzi generali e criticità; 6.- L'opzione antropocentrica; 7.- Conclusione provvisoria.

1.- Premessa

Il moltiplicarsi di moniti e l'agitarsi di preoccupazioni sugli impieghi dell'Intelligenza Artificiale (I.A.) nei diversi ambienti professionali, e gli scenari più o meno prevedibili aperti dai recenti approdi normativi eurounitari, in primo luogo il Regolamento UE 2024/1689 diretto a contrastare i rischi di tali impieghi, impongono di fermarsi a riflettere sulle ricadute della crescente introduzione delle tecnologie ICT nell'ambito del diritto ed in modo particolare in quelli della decisione giudiziale che vengono ad arricchirsi di prospettive sicuramente nuove rispetto a ruoli e impostazioni teoriche tradizionali.

È un dato di fatto la crescente potenza di calcolo dei sistemi di I.A. per automatizzare e implementare le performances dei servizi pubblici tramite algoritmi che “apprendono” dai dati. Ma ciò che dovrebbe indurre a riflettere è la crescita esponenziale delle immense riserve di dati a cui possono attingere gli algoritmi in vista di elaborazioni dirette a scopi diversi.

Si tratta di un processo che parte da lontano e in questi ultimi anni si è amplificato a dismisura espandendosi in ambiti ritenuti fino a poco tempo fa appannaggio esclusivo di inattingibili capacità umane. Non c'è lavoro, professione, attività sociale che possa chiamarsene fuori, men che meno le professioni legali continuamente chiamate in causa su aspetti tipici del ragionamento giuridico, dall'interpretazione dei testi normativi, alle attività professionali tipiche, con particolare attenzione alla decisione giudiziale.

Una rivoluzione epocale che chiama in causa aspetti diversi e molteplici criticità, tutte però riconducibili ad un nodo teorico principale: la possibilità della meccanizzazione del ragionamento giuridico o di suoi segmenti importanti.

Problema non nuovo dal punto di vista della logica, e della logica giuridica in particolare, perché discende direttamente dal noto *calculemus* seicentesco di leibniziana memoria¹ che ha trovato nelle recenti applicazioni dell'I.A. inaspettate conferme, fino a qualche decennio fa ritenute impensabili.

* Testo rivisto del seminario tenuto durante il corso di Logica Giuridica, a. a. 2023-2024.

** Professore aggregato di Logica Giuridica presso il Dipartimento di Scienze Giuridiche dell'Università degli Studi di Salerno.

¹ «Quando sorgeranno delle controversie, non ci sarà maggior bisogno di discussione tra due filosofi di quanto ce ne sia tra due calcolatori. Sarà sufficiente, infatti, che essi prendano la penna in mano, si siedano con una tabula, e si dicano

Le implicazioni di questo inarrestabile processo lasciano prevedere ulteriori imprevedibili sviluppi in tutti i campi, e non lasciano spazio all'esistenza di zone "franche".

Le previsioni dell'avvento di super-intelligenze in grado di oltrepassare le capacità cognitive naturali si presentano come una sorta di nietschiano "umano troppo umano", in grado di trasvalutare i valori correnti e di aprire a scenari che non mancheranno di far sentire presto i loro effetti sul secolo di cui abbiamo appena cominciato a svelare le enormi potenzialità.

Sul piano che più da vicino ci interessa, l'I.A. ha mosso passi significativi nel mondo giudiziario con la sperimentazione di strumenti per la cd. giustizia predittiva davanti ai quali ogni tentazione di conservazione dell'assetto attuale di poteri pubblici e privati appare a dir poco velleitario.

La legge francese sul divieto dei *tools* di giustizia predittiva come PREDICTICE, che consentono la profilazione dei giudici sulla base della raccolta delle loro decisioni e dei "bias" che essi contengono, è uno degli inequivocabili segnali dei rapidi e imprevedibili sviluppi delle tecnologie di I.A. applicate all'ambito giudiziario².

I sistemi sono già oggi in grado di analizzare una quantità straordinaria di dati storici e di fare previsioni con un livello di probabilità molto alto circa l'esito di giudizi e decisioni.

La sperimentazione avanzata di sistemi come Compas nel sistema giuridico americano, RoboJudge e AI Lawyer nel Regno Unito, e molti altri applicativi che vengono progettati e sperimentati di continuo in Europa, dimostrano non solo l'interesse del mercato verso questi prodotti ma le attese espresse da molte parti quanto all'efficientamento dei sistemi giudiziari, pur ponendo, del pari, seri problemi riguardo alla mancanza di trasparenza degli algoritmi che ne costituiscono l'ossatura logica.

2.- L'analisi predittiva sulla base dei dati

Sono molti gli aspetti che la ricerca sta affrontando in questo campo, in primo luogo sulle caratteristiche dell'analisi predittiva delle decisioni giudiziali, cioè sull'utilizzo di algoritmi per la meccanizzazione di un importante segmento del ragionamento giuridico, basato su *cluster* di dati storici (gruppi e sottogruppi).

Una varietà di strumenti già oggi presenti sul mercato offrono applicazioni interessanti alle prospettive aperte in questo settore che utilizza tecniche di "data mining" e di "machine learning", su basi essenzialmente statistiche, con lo scopo immediato di prevedere i possibili esiti di cause legali.

Ma al di là dell'evidente obiettivo pratico dell'impiego di questi strumenti destinati in primo luogo agli studi legali per prendere decisioni strategicamente informate, risulta interessante dal punto di vista teorico guardare alla costruzione di questi strumenti e, quindi, ai vantaggi ma anche alle criticità che essi nascondono.

reciprocamente (chiamato, se loro piace, un amico): calcoliamo (...): G.W Leibniz, *Scritti di logica*, a cura di F. Barone, Bologna 1968, 237.

² Ci si riferisce in modo particolare all'art. 33 della L. n. 222/2019 di programmazione 2018-22 per la riforma della Giustizia, con la quale è stata vietata nell'ordinamento francese la pubblicazione del nominativo degli autori delle sentenze allo scopo di evitare l'attività predittiva e la profilazione dei giudicanti e dei loro orientamenti, resa concretamente possibile oggi da "legal tools" come Predictice, il cui software di analisi consente di aumentare e rendere sempre più efficace la predizione di orientamenti di giudici e colleghi giudicanti. Va considerato che "tools" dello stesso tipo, sono già in uso nei paesi anglosassoni e offrono potenzialità analitiche non minori.

Non si tratta, come ingenuamente ritengono i poco informati, di una mera raccolta di dati storici relativi a casi legali del tipo risultati, giudici, avvocati, giurisdizioni e variabili contestuali. La presentazione del complesso delle fasi di processazione standard cui sono sottoposti i dati sta a dimostrarlo.

L'approccio fondamentale è svolto dal cosiddetto "Data Pre-Processing" che consiste nell'operazione di ripulitura e di trasformazione dei dati bruti.

Si tratta, innanzitutto, della rimozione di possibili duplicati ma non solo. La processazione logica dei dati comporta l'integrazione di valori assenti e l'eventuale normalizzazione dei campioni ai fini della loro classificazione e comprensione. Questo implica la creazione di variabili derivate, i cosiddetti "feature engineering" che rappresentano un aspetto determinante nel miglioramento delle performance di predittività del modello in uso.

Un secondo, ma non minore, aspetto riguarda la scelta stessa dell'algoritmo, perché da questo dipendono gli output concreti della ricerca.

Le tecniche di "machine learning" che possono essere implementate in vista dell'analisi predittiva costituiscono uno degli aspetti di maggiore criticità dal punto di vista della logica del sistema e dei risultati cui possono condurre³.

Se da una parte, infatti, la Regressione Logica rappresenta un modello di base utile per la previsione di esiti binari con la coppia vittoria/soccombenza, d'altra parte è possibile, e talora necessario, perfezionare il percorso con l'utilizzo di modelli logici complessi, destinati a classificare le decisioni basate su variabili multiple che vanno oltre la semplice coppia vittoria/soccombenza.

Si pensi per esempio agli Alberi Decisionali o, ancora, alle Reti Neurali per comprendere i processi decisionali di casi complessi con un maggiore e talora elevato livello di variabilità nei dati.

A questo scopo si comprende l'inevitabilità di un terzo step, fondamentale, nei processi svolti in sede di analisi predittiva dei dati e cioè un adeguato periodo di addestramento.

I modelli devono essere "addestrati" sui dati storici e l'algoritmo deve imparare a "riconoscere" gli schemi logici e le correlazioni che si ripetono nelle serie storiche somministrate.

Di solito si utilizza una porzione dei dati per l'addestramento e un'altra per la validazione del modello.

Ma ancora non basta perché, dopo l'addestramento i modelli per l'analisi predittiva dei dati giudiziari vanno ancora verificati sul campo, dal punto di vista cioè delle metriche di performance di cui essi sono capaci, e cioè in primo luogo la precisione con la quale il modello riesca a fare previsioni, per esempio, con riferimento ad uno strumento statistico di base come sono le cosiddette curve ROC ("Receiver operating characteristic") utilizzate in una grande varietà di campi scientifici, ma particolarmente usate nell'ambito della teoria delle decisioni per la valutazione dei classificatori

³ La profilazione è una delle tecniche più diffuse nel trattamento automatizzato dei dati, grazie ad essa è possibile creare conoscenza predittiva, correlando i dati e la costruzione di profili, cioè di pattern destinati ad essere impiegati nell'assunzione di decisioni. La scoperta di schemi all'interno di grandi insiemi di dati è ciò che identifica profili associabili a un soggetto singolo ma anche a gruppi di soggetti. Cfr. F. Bosco, N. Creemers, V. Ferraris, D. Guagnin, E.J. Koops, *Profiling Technologies and Fundamental Rights and Values: Regulatory Challenges and Perspectives from European Data Protection Authorities*, in *Reforming European Data Protection Law*, Dordrecht 2015, 3-33.

binari sulla base della relazione tra i tassi di veri positivi per il modello preso ad esame e i tassi di falsi positivi⁴.

Da ultimo, “last but not least”, dopo che il modello è stato convalidato, va sottoposto al “training” necessario perché sia applicato ai nuovi casi che vengono somministrati per ottenere previsioni sugli esiti: si tratta della parte più delicata che ha rilievo dal punto di vista della ricaduta pratica di questi strumenti e cioè l’utilizzo delle informazioni per migliorare la strategia in un caso concreto in cui si tratti di decidere per esempio sull’alternativa tra il procedere con un’azione legale o tentare una possibile negoziazione.

Sono evidenti i vantaggi discendenti dall’analisi predittiva dei dati nei diversi ambiti legali. L’ottimizzazione delle risorse, innanzitutto, perché gli studi legali possono concentrare gli sforzi sui casi con maggiori probabilità di successo e decidere in quali casi investire risorse, minimizzando i costi, pianificando le risorse, ripartendo in modo efficace compiti e orari di lavoro nella preparazione degli argomenti.

Ma al di là degli immediati vantaggi perfettamente intuibili, nell’uso di strumenti di questo tipo, si affacciano questioni teoriche che non possono essere eluse, quanto ai limiti ma anche alle potenzialità ancora inesplorate delle analisi basate su tali metodi; criticità che la ricerca di base è chiamata ad affrontare a partire dalle “logiche” intrinseche alla costruzione di tali modelli.

Il plurale è d’obbligo, trattandosi di una varietà di metodi che chiamano in causa non la logica classica, non solo quella almeno, ma le sue estensioni: a partire dalle logiche modali aleatiche che conducono direttamente alle logiche deontiche (KD, D₄, D₅), ma anche alle logiche epistemiche, a quelle temporali, per spaziare poi nelle logiche polivalenti, la logica *fuzzy* in primo luogo, e poi le logiche della rilevanza, le condizionali, le paracoerenti, le non monotone che costituiscono la base indispensabile per tentare di comprendere i processi di costruzione della meccanizzazione del ragionamento giuridico, e tentare se non di dirigere, almeno di comprendere le strategie, gli obiettivi e la direzione stessa dell’Intelligenza Artificiale applicata al mondo del diritto.

Tra le criticità più evidenti un posto di primo piano spetta oltre che alla quantità dei dati, alla loro qualità, perché la selezione dei dati storici può influenzare notevolmente le previsioni fatte dai modelli distorcendo o rafforzando gli scopi dell’analisi.

Le correlazioni tra i dati che un sistema è in grado di stabilire, di solito aggregate quantitativamente in punteggi, conferiscono una maggiore o una minore probabilità rispetto al comportamento atteso, quindi in termini di predizione favorevole o sfavorevole di esso.

Ma a questo punto diventa fondamentale discernere la qualità dei dati e per esempio la loro provenienza, si pensi al caso critico di un sistema giudiziario: «Nel valutare l’impiego di un sistema automatico per effettuare predizioni, bisogna distinguere se i dati contenuti nell’insieme di addestramento siano costituiti da scelte passate di esperti umani, o invece da eventi indipendenti da tali scelte (...) Considerazioni analoghe potrebbero farsi per un sistema destinato ad operare nell’ambito della giustizia, ad esempio, per determinare se concedere o meno la libertà vigilata. Il sistema potrebbe essere addestrato: (a) sulla base di un insieme di addestramento che associa richieste di libertà vigilata alle corrispondenti decisioni dei giudici o (b) sulla base di un insieme di

⁴ Dal punto di vista logico e giuridico una regola di decisione può considerarsi una funzione probabilistica e rimanda a logiche di tipo bayesiano. Mi permetto di rinviare per un orientamento generale su tali argomenti a C. de Giacomo, *Inferenze. Elementi di Logica*, Soveria Mannelli 2010, cap. V: *Logica della Probabilità*, 195-242.

addestramento che associa le stesse richieste di libertà vigilata al comportamento successivo dell'imputato (indicando se questi si sia comportato correttamente o, invece, si sia sottratto alla pena o abbia reiterato il reato)»⁵.

Così come, non può tacersi di uno degli aspetti che maggiormente oggi richiamano l'attenzione degli studiosi di questi argomenti: i cd. "bias".

Se i dati storici presentano "bias", ovvero pregiudizi di qualche tipo, che possono essere diretti o indiretti, le previsioni non possono che perpetuarli, almeno entro certi limiti.

E' ben vero che gli algoritmi di solito non propongono decisioni basate su palesi presupposti discriminatori, ma è altrettanto vero che esistono forme indirette di discriminazione legate comunque a caratteristiche specifiche, come il genere, la razza e l'appartenenza religiosa, che nel determinare previsioni apparentemente imparziali, possono integrare nell'algoritmo conseguenze pregiudizievoli per gli individui appartenenti a determinati gruppi senza una adeguata motivazione.

D'altra parte proprio su questi argomenti è stato evidenziato in letteratura che in molti casi le decisioni automatizzate, anche quelle contenenti la valutazione di singoli individui, possono essere utili a portare alla luce le cd. discriminazioni indirette che si annidano nelle decisioni umane sotto forma di pregiudizi e fallacie nel ragionamento.

In questo senso l'algoritmo può presentarsi più preciso e trasparente delle decisioni umane che spesso richiedono a chi ne è destinatario, di conferire un elevato tasso di "fiducia" al decisore, una sorta di inevitabile sacrificio che può rivelarsi eccessivo ai limiti del fideismo.

È stato scritto che gli algoritmi decisionali riducono il "rumore", cioè l'errore medio, contenuto nelle decisioni umane, il che spiega perché casi simili possano mettere capo a decisioni diametralmente opposte senza un valido motivo: «molti approcci meccanici, dalle regole più banali agli algoritmi più sofisticati e impenetrabili, possono surclassare i giudizi umani. Un motivo fondamentale alla base di questa prestazione superiore, anche se non l'unico, è che tutti gli approcci meccanici sono privi di rumore»⁶.

Ma da molti altri punti di vista si è ritenuto che i risultati decisionali degli algoritmi possano risultare migliori di quelli prodotti da esperti.

Pur con le dovute cautele va preso atto che la semplice ricognizione della complessa costruzione delle fasi di processazione dei dati, preordinate all'analisi predittiva delle decisioni giudiziali da parte di sistemi di I.A., sta a dimostrare il grado di avanzamento di questo settore della ricerca che chiama in causa direttamente, per superarli, i modelli tradizionali di analisi della decisione giudiziale.

L'analisi predittiva delle decisioni si sta rivelando uno strumento sempre più potente per il settore legale, con un grado di efficienza che al momento già raggiunge percentuali seriamente comparabili con le capacità di previsione umana e lascia immaginare nuovi imprevedibili progressi in termini di meccanizzazione del ragionamento giuridico e di possibili ricadute sul funzionamento di importanti settori dell'attività legale in generale e di quella giudiziaria in particolare.

⁵ G. Sartor, *L'intelligenza artificiale e il diritto*, Torino 2022, 50.

⁶ D. Kahneman, O. Sibony, C.R. Sustain, *Rumore. Un difetto del ragionamento umano*, Torino 2021, 140.

3.- Il caso Loomis e il sistema Compas

Gli strumenti predittivi delle decisioni giudiziali rappresentano solo una parte delle sfide che abbiamo davanti, ma l'utilizzo dei sistemi predittivi di I.A. nell'attività giudiziaria rivela risvolti non secondari anche se riguardato dal punto di vista delle applicazioni utilizzate dagli stessi decisori.

La prima criticità riguarda le possibili distorsioni cognitive che possono ricorrere, e di fatto ricorrono in modo più o meno evidente, nei giudizi su fatti e circostanze che un individuo o un algoritmo è chiamato a valutare.

L'esistenza di "bias" occultati all'interno del giudizio è una ricorrenza tutt'altro che rara. Meglio non illudersi, dai "bias" non ci si libera facilmente, e probabilmente neppure definitivamente; il sistema giudiziario non fa eccezione. Semmai ci si può difendere da questa eventualità, entro certi limiti, con il contributo degli strumenti offerti dalla logica.

Consideriamo quello che può essere definito un *leading case* delle discussioni in argomento: la sentenza del 2016 della Corte Suprema del Wisconsin (State Wisconsin v. Eric L. Loomis, 13 Luglio 2016) sull'appello di una persona condannata a sei anni di reclusione. Cosa ha a che fare questo caso con le questioni di cui stiamo trattando?

La pena a cui era stato condannato l'imputato era stata elaborata in accordo ai risultati di un software specifico, COMPAS, acronimo di "Correctional offender management profiling for alternative sanctions", in base ai quali il profilo somministrato era da considerarsi quello di un soggetto ad alto rischio di recidiva.

La persona in questione, dopo essere stata fermata dalla Polizia alla guida di un'auto precedentemente coinvolta in una sparatoria, era stata imputata per cinque diversi capi di accusa, tra i quali la guida di un veicolo senza consenso del proprietario e il possesso di diverse armi da fuoco, e per ciascuno di tali capi l'accusato risultava recidivo. Dopo l'ammissione di colpevolezza da parte dell'imputato e la richiesta di patteggiamento per le accuse meno gravi, la Corte emetteva un provvedimento per ottenere un PSI ("Presentence Investigation Report"), e cioè una relazione con i risultati delle indagini sulla storia personale dell'imputato, al fine di determinare la pena. Il PSI includeva l'elaborazione dei risultati sulla base di un software dedicato, il COMPAS, appunto.

In particolare, COMPAS è uno strumento utilizzato per valutare il rischio di recidiva riferendosi ad una serie di indicatori del tipo l'occupazione svolta, la disponibilità di un alloggio, l'uso abituale di sostanze stupefacenti ecc. All'algoritmo associato a tale strumento veniva chiesto di elaborare i dati così ottenuti, ed il risultato, in termini di rischio di recidiva, era determinato sulla base di una scala da 1 a 10, mediante la quale valutare la rimessione in libertà del soggetto sulla base della propensione alla commissione di nuovi reati da parte di soggetti appartenenti a gruppi con indicatori simili⁷.

⁷ L'impiego di questo algoritmo nel sistema giudiziario americano ha una storia che risale indietro nel tempo: «Since 1998, COMPAS has been used by law enforcement and judicial officers in multiple US states to shape decisions about how to treat the criminal accused and convicted. As of 2022, more than a million defendants have received COMPAS risk scores, and many of those scores have been used to determine whether particular defendants should be incarcerated, and (if so) what the nature and duration of incarceration and any post-release supervision should be»: T. Dancy, *Artificial Justice*, New York 2023, Introduction, *passim*.

Naturalmente, l'algoritmo non è in grado di stabilire il rischio di recidiva individuale ma la previsione è elaborata operando una comparazione tra informazioni sul singolo e informazioni relative a gruppi di individui con caratteristiche simili.

Nel caso Loomis, i risultati di Compas attestavano un alto livello di rischio in tutti gli ambiti di recidiva considerati (pregiudiziale, generale, e quella di atti violenti).

Sulla base dei risultati ottenuti, considerata la gravità del crimine, l'ammissione di colpevolezza e la condotta dell'imputato, il Tribunale determinava la pena e negava a Loomis la libertà vigilata.

La difesa proponeva istanza di revisione della decisione lamentando che i risultati ottenuti con l'impiego dell'algoritmo avrebbero violato il diritto ad un processo equo, ma il Tribunale rigettava l'istanza di revisione, sostenendo che la pena inflitta sarebbe stata la medesima, a prescindere dalla considerazione dei risultati dell'algoritmo in questione.

La Corte Suprema del Wisconsin, chiamata a sua volta a decidere sulla questione, dettava alcuni principi destinati a far discutere, tra i quali in primo luogo quello secondo il quale il software può essere impiegato nei giudizi di determinazione della pena ma il suo uso deve essere ristretto entro limiti precisi perché i punteggi di rischio assegnati dall'algoritmo non possono essere utilizzati come fattori determinanti e univoci per la decisione sull'opportunità di concessione della libertà vigilata.

Ma ciò che più conta, in relazione agli aspetti trattati in questa sede, riguarda il motivo sollevato dal condannato riguardo l'impossibilità di confutare in modo appropriato l'uso dei dati processati da COMPAS, perché la copertura del brevetto e del segreto industriale avrebbero impedito alla difesa di accedere alle informazioni ed ai meccanismi di elaborazione attuati dal software.

Su questo punto la Corte Suprema del Wisconsin negava il pregiudizio lamentato poiché, pur restando segreti i processi di funzionamento, il manuale di COMPAS spiegava in modo sufficiente che i punteggi sono basati in gran parte su dati statistici, come quelli sulla condotta criminale, e sono affidati ad un numero limitato di variabili dinamiche, del tipo l'uso di stupefacenti. Il software contiene, infatti, una lista di domande e risposte riguardanti fattori statistici, di cui l'imputato aveva piena facoltà di verificare l'accuratezza, eventualmente contestandoli con informazioni e dati contrastanti.

I giudici supremi, nel citare i risultati di alcuni test effettuati da altri Stati americani, concludevano che, seppure non perfetto, COMPAS rappresentava un mezzo di calcolo affidabile, pur dovendo essere costantemente aggiornato sulla base dei cambiamenti sociali. D'altra parte la stessa Corte prescriveva ai Tribunali circondariali un uso accorto e limitato dell'algoritmo insistendo sull'esercizio prevalente della discrezionalità dei giudici nel valutare in modo bilanciato la storia di ogni individuo sulla base di tutti i fattori a disposizione. Sotto tali premesse la Corte escludeva che l'uso di COMPAS avesse violato il diritto all'equo processo dell'imputato.

Se da una parte, quindi, può considerarsi corretto considerare i dati forniti dal software ai fini della decisione, sarebbe sicuramente illegittimo basare la sentenza esclusivamente su tali risultati, utilizzandoli come fattori determinanti della decisione.

Nel caso di specie, infatti, il cd. punteggio di rischio aveva rappresentato solo uno dei numerosi fattori considerati dal Tribunale nel determinare la pena e il fattore rischio nel *decisum* aveva

ricevuto un peso minimo, al punto da indurre lo stesso Tribunale a considerare che la decisione non sarebbe stata diversa in assenza dei dati forniti dall’algoritmo impiegato⁸.

La sentenza che abbiamo ricordato è importante sotto una pluralità di motivi che hanno sollevato dubbi e discussioni in merito ai limiti nell’utilizzo degli strumenti predittivi nell’ambito delle decisioni giudiziali⁹.

Il primo problema, dal punto di vista teorico, riguarda la conoscibilità della logica e dell’architettura del software impiegato. Ci si è chiesto, infatti: «se il sistema fosse in mano pubblica e la sua architettura fosse conosciuta e condivisa dalla magistratura e dall’avvocatura, sarebbe comunque da rifiutare?»¹⁰.

In questa prospettiva, se la preoccupazione principale nell’uso delle tecnologie di IA nel diritto è rappresentato dall’esistenza di possibili “bias”, la preventiva discussione tra gli addetti ai lavori e il monitoraggio di banche dati e algoritmi, potrebbero superare le reticenze che ostacolano la diffusione di questi strumenti nel sistema giudiziario.

Si consideri che COMPAS non costituiva già nel 2016 un caso isolato, perché altri algoritmi dello stesso genere avevano già varcato la soglia delle Corti Statunitensi e sono tuttora in uso a supporto delle decisioni “pre-trial” dei giudici su questioni come la carcerazione preventiva e il rilascio su cauzione, si pensi a PSA, acronimo di Public Safety Assessment, che in diversi Stati, come Arizona e New Jersey, opera sulla scorta di variabili del tipo il rischio di fuga, l’incapacità di pagare la cauzione ecc.

Del resto, in linea di principio, perché dovrebbe essere più affidabile il ragionamento giuridico umano, la cui opacità nel processo motivazionale di una sentenza potrebbe celare l’esistenza di “bias” molto più profondi e opachi di quanto non accada nel codice di un algoritmo che si affida espressamente a dati quantitativi?

La ormai celebre domanda “preferite essere giudicati da un robot o da un giudice umano?” posta al convegno organizzato dallo Stanford Center for Legal Informatics nel 2015 sul futuro del diritto, rimane inquietante sullo sfondo di tutti i dibattiti che hanno ad oggetto questi argomenti, e con essa la risposta, ancora più inquietante, che la maggioranza dei convegnisti diede in quell’occasione a tale interrogativo, preferendo il robot all’uomo¹¹.

4.- Gli sviluppi del dibattito

I sistemi di “machine learning” adoperati nei sistemi giudiziari attuali, sono basati su strumenti di “big data analysis”, vale a dire enormi masse di dati strutturati (sentenze in primo luogo, ma anche

⁸ Wisconsin Supreme Court, *State v. Loomis*, 881 NW 2d 749 (Wis 2016). Per le riflessioni in argomento, cfr. K. Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis* in *North Carolina Journal of Law & Technology*, vol. 18, 2016. Più in generale, cfr. D. Talia, *La società calcolabile e i Big Data: algoritmi e persone nel mondo digitale*, Soveria Mannelli 2018; L. D’Agostino, *Gli algoritmi predittivi per la commisurazione della pena* in *Diritto Penale Contemporaneo*, n. 2/2019; A. Jean, *Nel paese degli algoritmi*, Vicenza 2021.

⁹M. Gialuz, *Quando la giustizia penale incontra l’intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa* in *Diritto penale contemporaneo*, maggio 2019.

¹⁰ A. Santosuosso, *Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano 2020, 109.

¹¹ CodeX FutureLaw 2015, third annual conference, Stanford Law School, Stanford University, 30 aprile 2015.

atti defensionali, provvedimenti amministrativi ecc.) inseriti in un ambiente informatico (un server che può trovarsi anche lontano dall'ufficio di provenienza) nel quale i dati sono trasformati in un insieme di informazioni che possono essere aggregate e disaggregate per scopi diversi.

La massiccia dematerializzazione degli atti giudiziari ha fornito un impulso straordinario a questo processo che costituisce ormai un punto di non ritorno.

La quarta rivoluzione industriale, come è stata da più parti ribattezzata l'epoca che viviamo oggi¹², ci pone di fronte a un dilemma fondamentale a proposito delle tecnologie digitali, esse infatti «non sono solo il modo in cui interagiamo con il mondo, come la ruota o il motore. Sono soprattutto sistemi che danno forma (formattano) e influenzano sempre più il modo in cui comprendiamo il mondo»¹³.

In altre parole, siamo davanti a una vera e propria re-ontologizzazione del mondo. Le tecnologie di cui parliamo non si limitano ad aumentare le possibilità dei soggetti che le utilizzano per le proprie attività, ma perimetrano e, di più, costruiscono nuovi spazi per chi deve pensare, argomentare, difendere/difendersi, decidere.

Si pensi, come esempio, a quanto l'analisi dei "big data", con strumenti di "machine learning", consenta oggi non solo di consultare in tempo reale una quantità di dati straordinaria, un tempo inimmaginabile per un decisore umano o per un ricercatore, ma come questa analisi permetta di scoprire correlazioni non immediatamente evidenti ad un essere umano e come da questi "pattern" possano essere dedotte nuove informazioni. Su queste basi e con questi obiettivi vengono progettati i sistemi di decisione automatizzata.

Il mondo giuridico può chiamarsi fuori da questi processi? Sembrerebbe proprio di no. Nonostante gli scetticismi più o meno diffusi e le reazioni più o meno gelose dei poteri tradizionali, il mondo del diritto è stato eletto a campo di ricerca per i "big data".

Il ragionamento giuridico dal punto di vista della logica è un ragionamento sotto vincoli. Tali vincoli, nel caso del diritto, sono rappresentati in primo luogo dalle regole che dirigono le decisioni.

Ma, come è stato scritto in modo illuminante, le decisioni guidate da regole in senso stretto costituiscono solo una parte delle decisioni giudiziarie: «rule-governed decision-making is a subset of legal decision-making rather than being congruent with it»¹⁴.

Un esempio emblematico è rappresentato dai casi tutt'altro che infrequenti di decisioni paramtrate su clausole generali, massime di esperienza, indicatori del tipo "nel migliore interesse del minore", nella scelta tra massimi e minimi della pena, del risarcimento, ecc.

Semmai il problema principale risulta essere quello della coerenza/congruenza delle decisioni rispetto ai principi del sistema; non a caso, quanto all'uso dell'intelligenza artificiale nei sistemi giudiziari, uno dei primi e fondamentali punti di riferimento è costituito a livello eurounitario dalla Carta Etica degli Usi dell'Intelligenza artificiale nei sistemi giudiziari imperniata, com'è noto, non su norme ma su cinque fondamentali principi¹⁵.

¹² L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano 2017.

¹³ L. Floridi, *The Logic of Information*, Oxford 2019, tr. it. Id., *Pensare l'infosfera, la filosofia come design concettuale*, Milano 2020, 24.

¹⁴ F. Shauer, *Playing by the rules: a Philosophical Examination of Rule-Based Decision-Making in Law and Life*, Oxford 1991, 11.

¹⁵ *European Ethical Charter of the Use of Artificial Intelligence in Judicial Systems and their environment* (Strasbourg 3-4 December 2018), documento adottato dalla Commissione del Consiglio d'Europa per l'efficienza della giustizia –

Su tali presupposti appare sempre più verosimile agli studiosi e agli operatori del diritto l'inevitabilità del passaggio da sistemi giudiziari integralmente "rule-based" a sistemi "data-driven", dove il riferimento ai "big data" e agli strumenti di "machine learning" diventano presupposti concettuali e funzionali ineludibili con i quali oggi si devono fare i conti.

Non che le norme manchino in questo ambito ma il loro impiego sembra piuttosto diretto, in questo particolare momento storico, a tentare di stabilire cippi confinari su terreni dall'andamento incerto o sollevare "warnings" di vario genere per i poteri pubblici e gli operatori privati.

Si pensi principalmente al Regolamento UE n. 679 del 2016 in materia di trattamento dei dati personali e di privacy che ha tra i propri obiettivi principali quello di rafforzare il controllo da parte dei cittadini europei sull'utilizzo dei dati personali da parte dei titolari del trattamento dei dati. Problematica che ha una ricaduta non secondaria nelle questioni che stiamo affrontando. Non a caso, agli artt. 13 e 15 del GDPR, come contraltare alla trasparenza algoritmica, è riconosciuto il diritto dell'interessato a conoscere l'esistenza dei processi decisionali automatizzati che lo riguardano, e agli artt. 22 e 71 dello stesso testo viene riconosciuto il diritto di contestare i processi decisionali automatizzati, ovvero i trattamenti discriminatori che siano frutto dell'esecuzione degli algoritmi. Questioni che pongono delicati problemi di bilanciamento collocandosi spesso all'incrocio tra la tutela del dato personale e la titolarità e l'utilizzo di banche dati di capacità straordinarie, in cui le informazioni affluiscono da fonti diverse e talora con dubbi processi di anonimizzazione.

Ma i problemi della governance istituzionale di questi processi hanno rappresentato uno scoglio ineludibile già in fase di proposte e di lavori preparatori dell'attuale assetto normativo eurounitario¹⁶. Dal canto suo, anche la giurisprudenza che si è fatta faticosamente largo su questi terreni scivolosi è stata costretta a prendere atto della mobilità dei confini di cui si discute.

La sentenza n. 2270/2019 del Consiglio di Stato, è stata una delle prime decisioni giudiziarie nel nostro Paese che, a proposito di applicazioni dell'I.A. nella Pubblica Amministrazione, ha tentato di fissare alcuni paletti sull'utilizzo di strumenti automatizzati nell'assunzione di decisioni da parte di soggetti pubblici.

Nel caso sottoposto a giudizio, relativo alla legittimità delle graduatorie degli insegnanti di scuola secondaria, essendosi accertato che il procedimento era stato rimesso integralmente all'esecuzione di un algoritmo, senza l'intervento di un funzionario umano nel controllo delle singole decisioni, la Corte decideva per l'accoglimento del ricorso degli insegnanti, ritenendo violato l'art. 97 Cost. e più in particolare i principi di imparzialità, pubblicità e trasparenza della decisione, considerando che non fosse stato possibile comprendere le modalità e le ragioni delle assegnazioni disposte dall'algoritmo. Anche in questa occasione, tuttavia, il supremo consesso della giurisprudenza amministrativa riconosceva l'indispensabile e vantaggioso utilizzo dell'I.A. nei processi decisionali amministrativi standardizzati ad esclusione dei casi di esercizio della discrezionalità, nei quali la valutazione degli interessi in gioco non può essere rimessa, non integralmente almeno, all'esecuzione di un algoritmo.

CEPEJ: principio del rispetto dei diritti fondamentali; principio di non-discriminazione; principio di qualità e sicurezza; principio di trasparenza, imparzialità ed equità; principio del controllo da parte dell'utilizzatore.

¹⁶ S. Pelleriti, *Breve analisi della governance istituzionale delineata nella proposta di Regolamento UE sull'intelligenza artificiale*, in *Riv. trim. dir. economia*, supplemento al n. 3/2021, 280 ss.

A fronte della perdurante incertezza negli ambiti di cui parliamo ha visto la luce, dopo lunghi anni di dibattiti sull'argomento, il recentissimo Regolamento Europeo 2024/1689, entrato in vigore il 2 agosto 2024, (ma la maggior parte delle norme inizieranno ad applicarsi tra il 2025 e il 2026) col quale il Parlamento Europeo ha elaborato un pacchetto di norme armonizzate per il territorio dell'UE, volte in primo luogo a chiarire gli obblighi degli sviluppatori, degli operatori e degli utilizzatori dell'I.A. in relazione agli usi che di quest'ultima viene fatta.

Considerata la progressiva diffusione di questi strumenti nei processi istituzionali, come nella vita delle imprese, delle famiglie e dei singoli individui, il testo di cui parliamo rappresenta il primo pionieristico tentativo di dare una risposta globale ai rischi di effetti indesiderati derivanti dall'aumento nella potenza di calcolo dei sistemi di I.A. e dal possibile impatto negativo sui diritti fondamentali.

5.- Il Regolamento EU 2024/1689 tra indirizzi generali e criticità

La prima legge sull'Intelligenza Artificiale, nota come A.I. Act, costituita da un Preambolo di 180 considerando, 113 norme e 13 allegati, è centrata su un concetto fondamentale, quello di rischio.

L'architettura "risk-based" del Regolamento proviene dai lavori preliminari che hanno preceduto il testo definitivo ed in particolare dalle linee definite nel Libro bianco sull'intelligenza artificiale, pubblicato nel 2020 dalla Commissione Europea alle quali il Parlamento si è evidentemente ispirato. Già il Libro bianco, infatti, proponeva un approccio fondato sui rischi, con particolare attenzione ai sistemi ad alto rischio, cioè quelli dell'I.A. in uso nei settori ritenuti critici, com'è quello di cui ci occupiamo in questa sede.

Il Regolamento ha tradotto molte delle raccomandazioni contenute nel Libro Bianco in obblighi legali vincolanti per i soggetti interessati.

In particolare, con riferimento ai temi trattati nel presente lavoro, il Libro Bianco ammoniva: «l' IA può svolgere molte funzioni che in precedenza potevano essere esercitate solo dagli esseri umani. I cittadini e le persone giuridiche saranno pertanto sempre più soggette ad azioni realizzate e a decisioni prese da sistemi di IA o con la loro assistenza».

A fronte dell'aumento delle possibilità di controllare le abitudini quotidiane delle persone, il Libro metteva in guardia dal «rischio potenziale che l'I.A. venga utilizzata, in violazione della normativa dell'UE sulla protezione dei dati e di altre norme, dalle autorità statali o da altri soggetti a fini di sorveglianza di massa» permettendo di analizzare grandi quantità di dati e di individuare collegamenti tra di essi, il che implica «nuovi rischi per la protezione dei dati personali anche in relazione a set di dati che di per sé non contengono dati personali»¹⁷.

Quanto al rischio dei "bias" nelle decisioni, il Libro Bianco ricordava: «Le distorsioni e le discriminazioni rappresentano un rischio intrinseco di qualunque attività sociale od economica. Il processo decisionale umano non è immune da errori e distorsioni. Queste stesse distorsioni, se presenti nell'IA, potrebbero tuttavia avere effetti molto maggiori e colpire o discriminare numerose persone in assenza dei meccanismi di controllo sociale che disciplinano il comportamento umano».

¹⁷ *Libro Bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*, Commissione Europea, Bruxelles 2020, 12.

La distorsioni che possono trovarsi nei sistemi di I.A. non è detto che siano frutto di errori di progettazione ma potrebbero piuttosto essere conseguenza di quanto il sistema ha “appreso” nel corso del suo utilizzo ed in particolare degli effetti delle operazioni di correlazione che il sistema è in grado di individuare all'interno di ampi “set” di dati.

Per tale motivo la Commissione avvertiva: «Le caratteristiche specifiche di molte tecnologie di IA, tra cui l'opacità (effetto “scatola nera”), la complessità, l'imprevedibilità e un comportamento parzialmente autonomo, possono rendere difficile verificare il rispetto delle normative dell'UE in vigore volte a proteggere i diritti fondamentali e possono ostacolarne l'applicazione effettiva. Le autorità preposte all'applicazione della legge e le persone interessate potrebbero non disporre dei mezzi per verificare come sia stata presa una determinata decisione con il coinvolgimento di sistemi di IA e, di conseguenza, se sia stata rispettata la normativa pertinente. Le persone fisiche e giuridiche possono incontrare difficoltà nell'accesso effettivo alla giustizia in situazioni in cui tali decisioni possono avere ripercussioni negative su di loro»¹⁸.

Sono quattro i livelli di rischio dei sistemi di I.A. definiti dalla normativa in ordine progressivo di minaccia alla sicurezza e vanno dal rischio minimo, al rischio limitato, fino al rischio elevato ed infine, al vertice della piramide, al rischio inaccettabile.

Escluso l'ultimo livello che riguarda le pratiche di intelligenza artificiale vietate (la cui applicazione è stata rinviata al 2 febbraio 2025), l'utilizzo di sistemi di I.A. nell'amministrazione della giustizia è associato ai rischi definiti come elevati.

In estrema sintesi potremmo dire che il combinato disposto degli artt. 9 e 14 del Regolamento, impone la sorveglianza umana e il controllo nei sistemi di IA ad alto rischio.

In relazione a tale esigenza è stato previsto che innanzitutto la progettazione debba essere tale da garantire la supervisione di persone fisiche durante l'intera durata del loro utilizzo. Ma altri obblighi più pervasivi, per progettisti e utilizzatori sono associati all'utilizzo di questi sistemi di I.A. tra i quali la previsione di sistemi di mitigazione dei rischi, l'analisi rigorosa della qualità dei dati che alimentano il sistema per ridurre al minimo i risultati discriminatori, la tracciabilità dei risultati e la trasparenza sulle finalità del sistema stesso perché se ne possa apprezzare il grado di conformità alla tutela dei diritti degli interessati.

6.- L'opzione antropocentrica

La ricostruzione di questo passaggio normativo e dei suoi precedenti, pur sommariamente accennati, consente di svolgere alcune considerazioni in margine alle premesse di natura teorica dell'A.I. Act, che non mancheranno di costituire oggetto di interrogativi e discussione in sede di applicazione del Regolamento.

Va considerata, soprattutto, la portata “antropocentrica” espressamente dichiarata all'art. 1 del Regolamento, a mente della quale costituirebbe preciso scopo dell'A.I. Act quello di «promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile».

Non appare chiara l'accezione del termine che il legislatore ha inteso utilizzare per definire l'obiettivo di questo documento normativo in vista dell'utilizzo appropriato dell'intelligenza

¹⁸ *Libro bianco*, cit., 13.

artificiale: se cioè esso sia da intendere come vagamente riferito alla centralità dell'essere umano, quindi come una mera clausola di stile, ovvero, in modo più penetrante, sia commisurato all'altro termine utilizzato: l'"affidabilità" dei sistemi di I.A.

In questo secondo caso, tuttavia, non è chiaro entro che limiti debba attendersi che un sistema di I.A. sia affidabile.

Questo aspetto non è necessariamente sovrapponibile (o realizzabile) con l'integrale trasparenza di tutte le fasi dell'elaborazione algoritmica, o di inferenze e correlazioni che i sistemi di intelligenza artificiale possono realizzare.

Del resto, tanto non viene comunemente richiesto alle decisioni umane, nei diversi contesti in cui esse si esplicano e producono effetti. In primo luogo nell'ambito giudiziario dove è accettato come normale un certo grado di "opacità" delle decisioni, essendo prescritto che la motivazione debba essere "concisa": si pensi significativamente nell'ordinamento italiano agli artt. 132 e 134 del codice di procedura civile, ovvero all'art. 546 del codice di procedura penale, o ancora all'art. 88 del codice del processo amministrativo. Sappiamo quante oscurità possano celarsi dietro la cosiddetta concisione.

La totale trasparenza, giustamente sacrificata sull'altare dell'economia e rapidità dei servizi resi, come quello della Giustizia, fino a che punto può considerarsi preclusiva di decisioni basate su sistemi di I.A.?

Per quale motivo, un certo grado di "opacità" delle decisioni basate su algoritmi operanti su grandi masse di dati dovrebbe essere di ostacolo ai vantaggi indubitabili che ne possono derivare quanto all'economia e alla rapidità della funzione? Questione non diversa, evidentemente, per quegli ambiti tradizionalmente riconosciuti come seriali dagli stessi operatori della Giustizia.

Ma c'è una seconda questione che s'impone all'attenzione, e non seconda per importanza.

Fino a che punto un operatore umano chiamato a "controllare" il funzionamento di siffatti algoritmi, la cui complessità è sicuramente impari rispetto alle capacità umane di elaborazione, può garantire che il controllo non sia puramente formale e che lo strumento di intelligenza artificiale utilizzato nella decisione sia realmente affidabile?

I cambi di paradigma in atto nell'I.A. sono una realtà inaggirabile: sembra superato il tempo nel quale operavano i primi sistemi esperti in cui gli operatori davano in pasto alle macchine delle regole per prendere decisioni del tipo "se A allora B", quello che in logica classica è il più comune condizionale materiale rappresentato da una freccia e da una tabella di verità.

Più che la centralità delle regole, nell'apprendimento automatico dei "machine learning", il centro dell'attenzione è puntato sui dati, e ancora meglio sui vasti contenuti di informazioni dai quali le macchine apprendono. La potenza di questo apprendimento risiede nelle cosiddette reti neurali, cioè nei codici composti da nodi/neuroni raggruppati in strati che formano la rete, le cui dimensioni multilivello dipendono dalla complessità dei compiti da svolgere, in cui ogni strato della rete neurale trasmette le informazioni a quello successivo, e la forza e i pesi delle connessioni sono variabili.

Ora, le reti neurali impiegano oggi architetture logiche di apprendimento cosiddetto "supervisionato" o "non-supervisionato" per stabilire la forza delle connessioni, e al posto del ragionamento passo-passo dei sistemi esperti, della logica classica e al programma di ricerca di Leibniz, utilizzano sistemi di apprendimento dipendenti dai dati di addestramento. Maggiori sono i dati, maggiore l'efficienza del sistema rispetto alle decisioni da assumere. Ciò che i sistemi devono sapere non dipende in

ultima analisi dalle istruzioni o dalle regole dettate dal progettista, ma dai “pattern” che il sistema ricava dai dati di addestramento.

Dai campioni di dati forniti (e nel caso dei sistemi non-supervisionati da quelli non etichettati con i quali la macchina si confronta autonomamente), la rete neurale elabora modelli che confronta con le risposte/decisioni umane, e trovando degli scostamenti rimodula progressivamente connessioni e schemi diminuendo la possibilità di distinguere la risposta umana da quella cosiddetta “artificiale”¹⁹. E questo processo può riguardare tanto le reti che devono riconoscere per esempio i volti in un aeroporto a fini di sicurezza, come quelle reti che devono riconoscere o predire schemi di decisioni su fattispecie giuridiche.

Una volta che la rete ha “imparato” a ragionare sarà destinata a lavorare sui nuovi dati che le verranno forniti, con un grado di affidabilità pari e indistinguibile da quello umano, ma molto superiore a quello del singolo operatore umano in termini di expertise dei dati consultati, perché la potenza elaborativa di un sistema di “machine learning” è legata a immense banche dati consultabili in un tempo che a un uomo costerebbero anni e forse una vita.

Quale tipo di controllo da parte dell’operatore umano è concretamente possibile immaginare?

7.- Conclusioni provvisorie

Considerati i contenuti discussi finora e l’incertezza dei confini di cui abbiamo preso atto, l’unico tipo di conclusione che può essere verosimilmente azzardata rispetto ai vettori di ricerca in atto non può che qualificarsi come provvisoria.

Gli ambiti applicativi dell’I.A. che abbiamo preso in esame sono di quelli che il Regolamento UE cataloga nell’alveo dei rischi ammessi, pur con la consapevolezza che di rischio elevato si tratti.

Ora, a prescindere da ogni proclama antropocentrico, non sarebbe ragionevole, a giudizio di chi scrive, immaginare che l’ A.I. Act si trasformi in un sorta di miracoloso “aperti sesamo” di tutte le “black box” dei sistemi di “machine learning” di ultima generazione, impiegati negli ambiti più diversi e in quelli giudiziari ed amministrativi in particolare.

Sarebbe erroneo caricare di aspettative poco realistiche un testo normativo destinato in primo luogo a decisori umani e a funzionari responsabili di settori chiamati in modo crescente ad impiegare nel proprio lavoro algoritmi e “machine learning” operanti su “big data”. Così come sarebbe anacronistico ed antistorico immaginare di poter ignorare sistemi in grado di migliorare sensibilmente l’efficienza di servizi tradizionalmente affetti da ritardi e criticità e quindi in conflitto

¹⁹ Singolare analogia con un celebre “denken experiment” ideato nei lontani anni ’50 dello scorso secolo dal logico inglese A. Turing, con il quale si propone a una persona di tentare di capire attraverso le risposte fornite, se il soggetto ignoto che si trova in un’altra stanza, e al quale può essere rivolta ogni sorta di domanda per iscritto, sia realmente un umano o piuttosto una macchina. Se la macchina riesce ad ingannare l’interrogante o se quest’ultimo non riesce più a distinguere tra un umano e un computer, il test si considera superato: A. Turing, *Computing Machinery and Intelligence*, in *Mind*, LIX, 1950, 433-460. Per quanto il test non sia mai stato superato al 100%, vi sono recenti applicazioni di I.A. che stanno migliorando progressivamente i programmi in grado di ingannare una percentuale sempre maggiore di umani, tra le quali alcune versioni di Chat GPT-4.

con le esigenze del mercato e dell'utenza. Da molte parti viene ripetuto che non è credibile immaginare di “staccare la spina” delle nuove tecnologie.

Il problema, semmai, sarà di stabilire concretamente i termini futuri dell'interazione tra uomo e macchina, tenendo presente la fluidità dei confini fra gli agenti in gioco, come evidenziato in modo emblematico dalla compresenza tra gli algoritmi di “machine learning”, oggi disponibili sul mercato, di algoritmi basati sull'apprendimento supervisionato accanto a quelli non supervisionati.

Stiamo parlando di tecniche di addestramento degli algoritmi che non si basano più soltanto sui dati etichettati dall'uomo, quindi specificati sia in input che in output, ma anche su dati non etichettati dove cioè gli algoritmi possono apprendere direttamente da input sconosciuti con i quali entrano in contatto, possono autonomamente stabilire relazioni tra di essi, identificare modelli nascosti negli stessi e creare inferenze originali.

Si pensi alle tecniche sempre più diffuse e raffinate consistenti nel raggruppamento di dati non categorizzati in “cluster” sulla base di somiglianze nelle loro caratteristiche, ciò che consente di segmentare ogni sorta di gruppi e stabilire relazioni non visibili esplicitamente, affinando progressivamente le capacità del sistema²⁰.

Per avere un'idea della loro potenza si pensi alle applicazioni in materia di sicurezza che consentono di monitorare il traffico di una rete per predire possibili incidenti e prevenirne gli effetti. Di norma i valori considerati non sono individuati in modo deterministico ma si trovano in “range” di percentuali di probabilità rispetto a intervalli ritenuti normali. Anzi, spesso l'apprendimento non supervisionato è utilizzato proprio in funzione esplorativa, alla ricerca di possibili modelli da rinvenire nell'analisi di grandi banche dati, allo scopo di individuare possibili anomalie.

Indicatori e medie relativi alla densità del traffico o alla temperatura ambientale, pur essendo diversi dal punto di vista qualitativo, basano in modo efficiente l'operatività degli algoritmi sugli stessi metodi richiesti per stabilire l'affidabilità di chi fa domanda per un prestito o per la ricerca di anomalie nei movimenti internazionali di capitali, così come per inferire la probabilità della reiterazione di condotte criminali o per individuare operazioni rivelatrici di possibili evasioni fiscali, e via dicendo.

Una difesa realistica, non puramente di stile, dell'antropocentrismo, richiederebbe rispondere con onestà intellettuale a questa domanda: tenuto conto degli innegabili vantaggi rappresentati dalla qualità e dall'ottimizzazione dei tempi nella resa dei servizi, e considerata la minimizzazione dei costi degli stessi, a quanta parte dell'efficacia di questi strumenti i soggetti pubblici e privati sarebbero disposti a rinunciare pur di non sacrificare la trasparenza e la controllabilità da parte dell'operatore umano nella processazione dei dati compiuta autonomamente dal sistema?

I processi culturali e naturali in atto sembrano condurre, nella loro globalità, verso dimensioni oltre-umane, a dispetto di ogni proclamazione di sacri principi; una condizione in cui l'uomo ha una posizione decisamente decentrata rispetto alle epoche della storia che ci hanno preceduto.

²⁰ «Tutto quello che percepiamo, dai volti degli amici ai suoni delle parole, è un cluster di qualcosa. Se non si fossero i cluster saremmo perduti (...) Avendo a che fare con i big data -un numero molto grande di oggetti- la nostra prima reazione è di raggrupparli in un numero di cluster più facile da gestire. (...) Il mondo, tuttavia, è ben più ricco: all'interno di un cluster si annidano altri cluster, e così via fino ai singoli oggetti», P. Domingos, *L'algoritmo definitivo*, Torino 2016, 242 ss.

Il post-umanesimo, con i suoi approcci “more than human”, ci indica una strada nella quale l’uomo avverte di essere solo parte di un sistema ben più complesso, e di questo sistema le intelligenze artificiali fanno parte a loro volta per la soluzione di problemi altrimenti irrisolvibili. L’intelligenza non è una prerogativa degli esseri viventi e l’essere umano non ne è l’unica forma. Come è stato acutamente osservato: «Certo non abbiamo bisogno di presupporre che l’agente abbia un cervello, un linguaggio o una coscienza. Dobbiamo tuttavia ipotizzare che l’agente abbia degli obiettivi, in modo tale da poter definire come “efficaci” o “appropriate” quelle azioni che aumentano le sue possibilità di raggiungerli. L’idea di comportamento teso a un obiettivo è la base della cibernetica, che fu il primo tentativo di sviluppare una teoria unificata degli agenti intelligenti»²¹.

L’irresistibile ascesa dell’intelligenza non-naturale, quindi non-biologica, in ogni ambito della vita sociale, è una delle conferme più illuminanti del post-umanesimo, una tendenza in grado di resistere ad ogni aprioristica attestazione antropocentrica, come quella proclamata nell’A.I. Act, per quanto questa considerazione possa dispiacere o impensierire sotto diversi aspetti. D’altra parte, alla ricerca, quando è autentica, viene chiesto di non piangere e di non ridere (anche perché, diversamente, i ricercatori sarebbero distratti da troppe occasioni prossime per fare l’uno e l’altro).

I dati raccolti nei diversi settori interessati mostrano che l’apprendimento automatico dei sistemi di I.A. è in rapida ascesa e che l’interazione uomo-macchina tenderà in modo inarrestabile a pervadere tutti gli ambiti nei quali si prendono decisioni, da quelli produttivi e commerciali a quelli giudiziari e amministrativi.

Come si è già accennato, le tecniche più recenti utilizzate nella progettazione dei sistemi di I.A. ricorrono sempre più spesso alla cosiddetta NSA (“Neural Architecture Search”) nella quale i sistemi di apprendimento automatico progettano altri sistemi di apprendimento automatico. In termini diversi, essi sono rappresentati come scatole nere all’interno di altre scatole nere, che se da una parte migliorano la capacità dei sistemi e diminuiscono sensibilmente il tempo per la risoluzione dei problemi più complicati, dall’altro, inequivocabilmente e oggettivamente, riducono la possibilità per i tecnici di comprendere l’effettivo funzionamento del sistema stesso. Ciò che lascia predire agli esperti del settore un’evoluzione tecnica nella quale l’apprendimento automatico è destinato ad aumentare e non a diminuire la propria opacità²².

Resisterà un “diritto alla spiegazione” di fronte alla constatazione che sistemi alternativi a quelli umani sono più rapidi ed efficienti nella soluzione di problemi anche se meno trasparenti? E quale spiegazione potrà considerarsi sufficiente?²³

Il cosiddetto standard debole dell’I.A., rappresentato nei termini della possibilità del controllo permanente sulla macchina da parte dell’utilizzatore, viene troppo spesso idealizzato come una sorta di panacea miracolosa contro i rischi paventati o reali degli sviluppi tecnologici in corso ed in particolare del massivo impiego di algoritmi e “machine learning” nei processi gestionali e

²¹ N. Cristianini, *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna 2023, 14.

²² Cfr. B. Bucanhan, A. Imbrie, *Il nuovo fuoco. Guerra pace e democrazia nell’era dell’intelligenza artificiale*, Milano 2024, in part. cap. 4 sull’opacità dei sistemi; T. Elsken, J.H. Metzen, F. Hutter, *Neural Architecture Search: A Survey*, in “Journal of Machine Learning Research”, 2018; B. Zoph, Q. V. Le, *Neural Architecture Search with Reinforcement Learning*, in 5th International Conference on Learning Representations, 2016.

²³ S. Wachter, B. Mittelstadt, L. Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, n. 2, 2017, 76-99.

decisionali da parte delle istituzioni e delle imprese, soprattutto quando esse siano alle prese con i “big data”. Ci si deve interrogare, tuttavia, su quanto realistica sia in tutti i contesti considerati l’opzione legata alla controllabilità.

Il nodo centrale e ineludibile resta quello del confronto tra le tecniche inferenziali del ragionamento umano e l’implementazione assicurata dall’utilizzo di modelli statistico-probabilistici esercitati su grandi masse di dati da parte dei cosiddetti software intelligenti.

Da più parti ormai si evidenzia la centralità di questo confronto e non necessariamente con epiloghi “antropocentrici”.

I riflessi dei molteplici impieghi dell’I.A. nelle società democratiche, sono un’occasione utile per ripensare agli inevitabili cambi di prospettiva riferibili al mondo del diritto (o del diritto-transumanista?) ed in particolare alla decisione giudiziale.

È stato scritto che: «è cresciuta nella società la legittima pretesa che il giudice, nell’esercizio dell’arte del giudicare e nella pratica giudiziaria, sia un buon ragioniere e un decisore di qualità. Sicché la professionalità, l’etica e l’implementazione del grado di expertise accumulata dal giudice nell’utilizzo delle tecniche inferenziali del ragionamento e nella verifica degli schemi statistico-probabilistici, acquisiti con l’ausilio della tecnologia digitale, di software informatici e algoritmi predittivi o con l’apporto della robotica e della logica dell’IA, potrebbero certamente contribuire a restituire al funzionamento della giustizia penale una più adeguata immagine di efficacia e qualità»²⁴.

Non sappiamo ancora per quanto tempo solo di “ausilio” si tratterà, ma al momento resta ineludibile la domanda sollevata nei precedenti paragrafi di questo lavoro, che vorremo declinare però in modo leggermente diverso, si potrebbe considerare la versione transumanista di quella proposta a suo tempo dalla Stanford Law School e potrebbe suonare così: con quale tipo di robot andrà sostituito il decisore umano?

²⁴ G. Canzio, *Intelligenza artificiale, algoritmi e giustizia penale*, intervento svolto nel webinar organizzato dalla Fondazione Leonardo – Civiltà delle Macchine su “Processo penale e Intelligenza Artificiale” (20 ottobre 2020).