



**ETICA ED INTELLIGENZA ARTIFICIALE: UN PRIMO COMMENTO  
AL DOCUMENTO DEL CEPEJ SU EUROPEAN ETHICAL CHARTER ON THE  
USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS AND THEIR  
ENVIRONMENT**

*Avv. Carlo Forte*

Nel 1942 Isac Asimov pubblicava il racconto “Circolo vizioso” dove di fronte al comportamento di un robot impazzito (Speedy), offre la sua visione delle tre leggi della robotica, ovvero:

*“1. Un robot non può recare danno a un essere umano, né può permettere che, a causa del proprio mancato intervento, un essere umano riceva un danno.*

*2. Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Prima Legge.*

*3. Un robot deve proteggere la propria esistenza, purché tale autodifesa non contrasti con la Prima o con la Seconda Legge.”*

Le stesse leggi che creano un corto circuito nel comportamento del robot, in quanto nello spasmodico tentativo di obbedire ad un ordine che gli è stato impartito (avvicinarsi ad un pozzo di selenio, in ottemperanza alla seconda legge), si allontana (per salvaguardare la propria esistenza in ottemperanza alla terza ed alla prima, in quanto autodistruggendosi al contatto con il selenio, non potrebbe evitare di produrre un danno all'uomo), per poi ricominciare a girare a vuoto non svolgendo più alcuna funzione utile all'uomo, minando la stessa sussistenza dei protagonisti (che si trovavano sul pianeta Mercurio).

È solo l'intervento dell'uomo a risolvere il danno ed accompagnare il robot verso la giusta soluzione.

L'etica dell'uomo si rivela fondamentale ma, allo stesso tempo, le Leggi di Asimov predicono un problema che si pone oggi nell'ambito dell'elaborazione di principi applicabili all'Intelligenza Artificiale. Sono senz'altro principi metagiuridici, che dovrebbero abbracciare un'etica generale, ma che allo stesso tempo interrogano il giurista quando debbano essere regolati e categorizzati nell'ambito della tutela di diritti fondamentali o di diversa natura. Ad esempio, in campo medico, la libertà della cura può scontrarsi contro il dovere di una macchina che, dovendo assistere il malato, dovrebbe sapere cosa fare se lo stesso malato le desse l'ordine di non assisterlo più. In tal senso, il consenso del paziente (Seconda legge di Asimov) potrebbe non essere compatibile con l'assenza di assistenza che arrecherebbe un danno alla sua salute (Prima legge di Asimov).

Immaginiamo, inoltre, una situazione in cui un robot sia alla guida di un'automobile: di fronte al rischio imminente di incidente che mette in pericolo in eguale misura il passeggero ed un passante, quale scelta dovrebbe fare? Il comportamento finale innesca degli elementi di responsabilità che riguardano i principi generali del diritto oltre ad altri aspetti del diritto materiale. Rinviando l'approfondimento sul tema della responsabilità al contributo di Carla Secchieri pubblicato su questo numero, qui è sufficiente rilevare che l'Etica che sorregge le scelte deve essere conforme ai i principi di diritto generalmente applicabili.

Nel dicembre scorso, La Commissione europea per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ, acronimo di European Commission for the Efficiency of Justice) ha deciso di adottare la prima Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari (*European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*), al fine di contribuire, nella consapevolezza del crescente utilizzo dell'IA nell'ambito dell'amministrazione della giustizia, a definire limiti etici all'uso della stessa IA, soprattutto con il fine di coniugarli nell'ambito di quelli che sono alla base della Carta dei Diritti dell'Uomo del Consiglio d'Europa e della Convenzione del Consiglio d'Europa sulla protezione dei dati personali.

Il documento è molto interessante ed ha una immediatezza sorprendente, in quanto individua 5 principi fondamentali attorno ai quali sviluppare gli intenti sopra descritti, usando una tecnica che guarda alla estensione del sistema delle tutele CEDU al campo dell'IA.

Il primo principio richiama **il rispetto dei diritti fondamentali**, e comporta che sia nella fase della progettazione che in quella dell'attuazione di strumenti e servizi di intelligenza artificiale usati nei sistemi

giudiziari, i diversi passaggi devono essere compatibili e rispettosi con i diritti fondamentali. Se, dunque, gli strumenti di intelligenza artificiale sono utilizzati per risolvere una controversia o come strumento di assistenza nel processo decisionale giudiziario o per fornire orientamenti al pubblico, è essenziale garantire che essi non pregiudichino le garanzie proprie del processo, quali appunto il diritto di accesso alla giustizia ed il diritto a un processo equo, il tutto nel quadro del generale rispetto dello Stato di Diritto e della separazione dei poteri.

Il secondo principio è la **non discriminazione**, che in via generale è un corollario del principio di eguaglianza e comporta la prevenzione ed il divieto di qualsiasi discriminazione tra individui o gruppi di individui. Nel campo dell'IA esso dovrebbe prevenire o vietare le discriminazioni attuate dalle stesse macchine, ad esempio prevedendo che il trattamento dei dati possa rivelare la discriminazione in relazione a persone singole o a gruppi; la macchina potrebbe anche avere un valore aggiunto in tal senso, perché confrontando molti altri dati, potrebbe prevenire diseguaglianze di trattamento. In definitiva, un siffatto processo tenderebbe a controllare che in qualsiasi fase non si applichino trattamenti differenziati o comunque non si riproducano o aggravino discriminazioni. Il documento sostiene che particolare attenzione deve essere prestata sia nella fase di sviluppo che in quella dell'attuazione, soprattutto quando il trattamento è direttamente o indirettamente basato su dati "sensibili". Tali dati possono includere informazioni relative a presunte origini razziali o etniche, origini socioeconomiche, opinioni politiche, convinzioni religiose o filosofiche, appartenenza sindacale, dati genetici, dati biometrici, dati sanitari o dati relativi alla vita o all'orientamento sessuale. Una volta individuata la discriminazione, occorrerebbe prendere in considerazione misure correttive per limitare o, se possibile, neutralizzare tali rischi e sensibilizzare le parti interessate.

In tale campo, il ricorso all'apprendimento automatico e progressivo della macchina, ad uno con analisi scientifiche multidisciplinari, potrebbe rivelarsi un utile strumento per combattere le discriminazioni.

Il terzo principio fa riferimento ai concetti di **qualità e sicurezza**, applicabili in particolare nel delicato campo del trattamento dei dati relativi alle decisioni giudiziarie. Il documento osserva che è necessario che la macchina utilizzi fonti certificate, da un lato, e che i dati siano concepiti e trattati in modo multidisciplinare, ed in un ambiente tecnologico sicuro, dall'altro. Ne consegue che i progettisti/programmatori dei modelli di apprendimento automatico dovrebbero necessariamente collaborare con dei professionisti del sistema giudiziario, ovvero con giudici, procuratori, ed avvocati, ma anche con altre categorie (multidisciplinarietà) come, in via esemplificativa, ricercatori od operatori nei settori del diritto e delle scienze sociali (ad esempio, mediatori, economisti, sociologi, filosofi, etc.); il documento preconizza la costituzione di veri e propri gruppi multidisciplinari che, in brevi cicli di progettazione, producano modelli funzionali, i quali siano costantemente monitorati e migliorati attraverso i riscontri nella fase dell'applicazione.

Come si diceva poc'anzi, i dati basati su decisioni giudiziarie inseriti in un software che alimenta un algoritmo di apprendimento automatico dovrebbero provenire da fonti certificate e non dovrebbero essere modificati fino a quando non siano stati effettivamente utilizzati nel meccanismo di apprendimento (ambiente sicuro). Infine, l'intero processo dovrebbe essere tracciabile per garantire che non vi siano state modifiche che possano alterare il contenuto o il significato della decisione in corso di elaborazione.

Il quarto principio indicato nel documento fa riferimento a 3 principi dei sistemi giuridici, ovvero **trasparenza, imparzialità ed equità**, coniugandoli in un unico concetto che in via generale comporta che i metodi di trattamento dei dati devono essere accessibili e comprensibili, con la previsione anche di audit esterni. Nel settore dell'IA tali principi comportano preliminarmente la risoluzione degli eventuali problemi relativi alla proprietà intellettuale, osservando che i diversi Stati membri dovrebbero trovare delle soluzioni equilibrate (NB: non si parla di armonizzazione perché la materia non è di competenza del Consiglio d'Europa).

Il documento osserva anche che quando vengono utilizzati strumenti che possono avere conseguenze legali o che possono incidere significativamente sulla vita delle persone, è necessario che tali principi siano sempre presenti nell'intera catena progettuale ed operativa. Nello specifico, la trasparenza è necessaria



almeno nel processo di progettazione. Si parla qui della possibile 'completa trasparenza tecnica', che a volte è limitata dalla protezione dei segreti commerciali (limiti della proprietà intellettuale).

L'imparzialità richiede l'assenza di pregiudizi. E l'equità – ma il documento cita anche l'integrità intellettuale - evidenzia che il meccanismo deve privilegiare gli interessi della giustizia.

Il sistema deve essere accessibile e quindi spiegato in un linguaggio chiaro e familiare (per descrivere come vengono prodotti i risultati) comunicando, ad esempio, la natura dei servizi offerti, gli strumenti sviluppati, le prestazioni e i rischi di errore. Dovrebbero essere previste Autorità indipendenti (o esperti indipendenti) per certificare e verificare i metodi di trattamento e fornire preventivamente pareri.

Il documento richiama in nota le osservazioni di uno studio (MSI-NET) del Consiglio d'Europa su "Algoritmi e diritti umani" che è interessante proprio per il problema dell'equilibrio della proprietà intellettuale: "La fornitura al pubblico di interi algoritmi o del codice software sottostante è una soluzione improbabile nell'attuale contesto, in quanto le aziende private considerano il loro algoritmo come una chiave di software protetto dalla proprietà intellettuale. Tuttavia, ci può essere la possibilità di richiedere che vengano fornite al pubblico sottoinsiemi chiave di informazioni sugli algoritmi che indichino, ad esempio, quali variabili sono in uso, per quali obiettivi gli algoritmi vengono ottimizzati, i dati di formazione, i valori medi e le deviazioni standard dei risultati prodotti, o la quantità e il tipo di dati elaborati dall'algoritmo".

Il quinto ed ultimo principio costituisce un elemento di chiusura, prevedendo che debba comunque essere rispettata la volontà dell'utente, elaborando così il **principio per cui il dato deve essere "sotto il controllo dell'utente"**. Il principio prevede, tra l'altro, che l'utente sia adeguatamente informato e possa avere il controllo delle proprie scelte. L'opzione scelta è quella di impedire un approccio prescrittivo e, al contempo, garantire che gli utenti siano attori informati e controllino le loro scelte, evidenziando come esse debbano essere autonome, con il divieto di limitare tale autonomia proprio attraverso l'uso di strumenti e servizi di intelligenza artificiale.

I professionisti del sistema giudiziario dovrebbero, in qualsiasi momento, essere in grado di riesaminare le decisioni giudiziarie e i dati utilizzati per produrre un risultato e continuare a non esserne necessariamente vincolati alla luce delle caratteristiche specifiche di quel particolare caso.

Il documento prevede che l'utente deve inoltre essere chiaramente informato di qualsiasi trattamento preliminare di un caso mediante intelligenza artificiale prima o durante un procedimento giudiziario e avere il diritto di opporsi, in modo che il suo caso possa essere trattato direttamente da un tribunale ai sensi dell'articolo 6 della CEDU (accesso alla giustizia, diritto ad un equo processo, diritto ad una difesa tecnica, etc.).

Il documento evidenzia come sia importante anche l'informazione, evidenziando l'importanza di programmi di alfabetizzazione informatica per gli utenti e dibattiti con la partecipazione di professionisti del sistema giudiziario.

Naturalmente, in questa breve presentazione non è possibile evidenziare tutti i passaggi del documento e quindi si rinvia alla sua lettura per potere apprezzare gli elementi metagiuridici, giuridici e le riflessioni multidisciplinari. Interessante è, in chiusura, notare alcuni richiami che, in appendice, sono evidenziati.

In primo luogo, si osserva che la trasformazione digitale sta producendo nelle nostre società un effetto disomogeneo sui sistemi giudiziari degli Stati membri del Consiglio d'Europa, per il diverso stato di avanzamento, per il diverso uso di applicazioni pratiche (sia in termini di tecnologia che di supporto legale), e anche per la ricerca di una gestione informatica efficace che non indica una unica soluzione, per il momento.

Se negli Stati Uniti, gli "avvocati robot" sono già al lavoro e sembrano conversare in linguaggio naturale con gli esseri umani, la tecnologia legale sta progettando sempre più nuovi servizi legali offerti ad avvocati ed altri prestatori di servizi legali, per consentire un accesso alla giustizia. Da questa apparente dicotomia emerge che nel futuro il ruolo dell'avvocato dovrà essere non quello di competere con le macchine, ma quello di governarle; per fare questo è necessaria una preparazione multidisciplinare che parta dalla riforma del sistema formativo universitario. L'elemento qualitativo dell'essere umano sarà, quindi, quello



del controllo sul lavoro che la macchina farà in modo molto veloce e potenzialmente efficace, facendo intravedere uno scenario in cui non è tanto la fatturazione oraria ad indicare la qualità dell'avvocato, quanto la sua capacità di governare il processo e di intervenire per orientare e correggere la macchina laddove si dimostri necessario. Questa conclusione, che è del sottoscritto e, dunque, passibile di critica, sembra conforme alle stesse conclusioni cui era giunto Asimov nel tentativo di elaborare le leggi della robotica.

Quanto al sistema giudiziario, il mercato sta dimostrando una corsa a costruire sistemi che mirino persino a prevedere le decisioni dei giudici, addirittura prevedendo il potenziale margine di errore. Sono i cd strumenti di "giustizia predittiva", anche se poi il documento evidenzia che tale denominazione non è corretta, in quanto una cosa sono le possibilità offerte dal mercato, anche a supporto del funzionamento del sistema giudiziario, altra cosa è l'amministrazione della giustizia. E quindi anche qui il richiamo finale del documento sembra essere quello di evidenziare che prescindendo dallo strumento, il miglior giudice deve essere non colui che non sbaglia, ma colui che fa la scelta giusta.