

La Convenzione quadro del Consiglio d'Europa sull'IA: verso una regolazione umano-centrica tra diritti fondamentali, democrazia e Stato di diritto

di Daniela Cardamone

La Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale afferma un modello di regolazione umano-centrico che riallinea l'innovazione tecnologica ai principi dei diritti umani, della democrazia e dello Stato di diritto. Di fronte ai rischi di una discrezionalità algoritmica opaca e depoliticizzata, la giurisdizione è chiamata a riaffermare il primato della legalità e la centralità della persona nell'era degli algoritmi.

1. La genesi della Convenzione e il suo fondamento assiologico / 2. Struttura e contenuti essenziali della Convenzione / 3. Governance dell'intelligenza artificiale e centralità della persona / 4. Discrezionalità algoritmica, depoliticizzazione del potere decisionale e ruolo della giurisdizione

1. La genesi della Convenzione e il suo fondamento assiologico

L'irruzione dell'intelligenza artificiale nella sfera giuridica, istituzionale e sociale costituisce uno degli snodi più delicati e al tempo stesso paradigmatici del rapporto tra diritto e tecnica nella contemporaneità. L'uso crescente di sistemi algoritmici decisionali, predittivi e generativi impone, infatti, di interrogarsi non solo sulla loro compatibilità con i diritti fondamentali, ma più profondamente sulla tenuta dei presupposti democratici e sulla stessa struttura dello Stato di diritto.

È in questa prospettiva che assume particolare rilievo l'intervento del Consiglio d'Europa, il quale – coerentemente con la propria missione statutaria – ha inteso affrontare la sfida posta dall'intelligenza artificiale non già come questione settoriale o tecni-

co-regolatoria, bensì come problematica sistemica, direttamente incidente sulla tutela dei diritti umani, sul corretto funzionamento dei processi democratici e sull'osservanza della *rule of law*.

Muovendo da tale consapevolezza, la Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto, adottata il 17 maggio 2024 dal Comitato dei ministri del Consiglio d'Europa e aperta alla firma il 5 settembre 2024 a Vilnius, si presenta come il primo tentativo di elaborare, in forma pattizia e giuridicamente vincolante, un quadro normativo sovranazionale idoneo a governare l'impatto dell'intelligenza artificiale sull'assetto valoriale degli ordinamenti democratici.

La premessa teorica da cui muove il testo convenzionale è che gli obblighi assunti dagli Stati parti della Convenzione europea dei diritti dell'uomo non conoscono attenuazioni o sospensioni in ragione del

progresso tecnologico. Al contrario, come chiarito già dalla raccomandazione CM/Rec(2020)¹ sugli impatti dei sistemi algoritmici sui diritti umani, la trasformazione digitale accresce la responsabilità degli Stati di assicurare che la progettazione, lo sviluppo, l'implementazione e l'uso dei sistemi di IA avvengano nel pieno rispetto dei diritti e delle libertà fondamentali, che restano universali, indivisibili, interdipendenti e interconnessi.

L'ambivalenza dell'intelligenza artificiale costituisce, del resto, il punto di partenza irrinunciabile di qualsiasi riflessione giuridicamente avvertita sul tema. Da un lato, essa è in grado di offrire opportunità senza precedenti per la promozione del benessere individuale e collettivo, per il rafforzamento della partecipazione democratica, per il miglioramento dell'efficienza delle istituzioni e persino per una più efficace tutela dei diritti fondamentali. Dall'altro lato, la medesima tecnologia può tradursi in uno strumento di compressione dell'autonomia individuale, di discriminazione sistemica, di sorveglianza pervasiva, di manipolazione dell'opinione pubblica e, non da ultimo, può comportare un rischio significativo di impatto negativo sul processo democratico e sull'esercizio dei relativi diritti umani.

La Convenzione quadro nasce precisamente dall'esigenza di governare tale ambivalenza, evitando tanto l'approccio tecnofobico quanto l'accettazione acritica dell'innovazione, e collocandosi in una posizione di equilibrio che consenta di valorizzare le potenzialità dell'intelligenza artificiale senza sacrificare i valori fondamentali su cui si fondano le democrazie europee e, più in generale, gli ordinamenti ispirati ai principi dello Stato di diritto.

Sotto il profilo della genesi, il testo convenzionale è il risultato di un processo negoziale particolarmente lungo e articolato, sviluppatosi nell'arco di oltre quattro anni sotto l'egida del *Committee on Artificial Intelligence* (CAI), organismo intergovernativo succeduto all'*Ad Hoc Committee on Artificial Intelligence* (CAHAI).

La composizione del CAI e le modalità di conduzione dei lavori meritano di essere sottolineate, poiché riflettono già sul piano procedurale l'approccio inclusivo e multilivello che caratterizza l'intera Convenzione. Accanto ai 46 Stati membri del Consiglio d'Europa, hanno infatti partecipato in qualità di osservatori numerosi Stati non membri – tra cui Stati Uniti, Canada, Giappone e Israele – nonché una pluralità di attori della società civile, del settore privato e del mondo accademico, chiamati a confrontarsi sui

rischi e sulle opportunità dell'intelligenza artificiale in una prospettiva autenticamente transnazionale. Anche l'Unione europea ha preso parte alle negoziazioni tramite la Commissione UE, in virtù della decisione 2022/2349 e includendo nella sua delegazione rappresentanti dell'Agenzia dell'Unione europea per i diritti fondamentali (FRA) e del Garante europeo della protezione dei dati (EDPS).

Tale impostazione consente di comprendere la scelta, non priva di rilievo politico-giuridico, di configurare la Convenzione come un "trattato aperto", suscettibile di adesione anche da parte di Stati estranei al Consiglio d'Europa e destinato, nelle intenzioni dei redattori, a costituire uno *standard* di riferimento globale. In questa dimensione si colloca, altresì, la previsione di un numero relativamente contenuto di ratifiche per l'entrata in vigore², soluzione che appare funzionale a garantire una tempestiva efficacia dello strumento convenzionale in un contesto tecnologico in rapido mutamento.

Quanto alla struttura e alla finalità, la Convenzione si qualifica espressamente come strumento "quadro", ossia come atto destinato a fissare principi generali comuni piuttosto che a dettare una disciplina di dettaglio. Essa non intende regolamentare l'intelligenza artificiale in quanto tale, né affrontarne gli aspetti economici, industriali o di mercato, che restano estranei al mandato del Consiglio d'Europa. L'obiettivo perseguito è, piuttosto, quello di garantire che gli obblighi già esistenti in materia di diritti umani, democrazia e Stato di diritto trovino piena applicazione anche rispetto alle attività rientranti nell'intero ciclo di vita dei sistemi di IA, dalla fase di progettazione e addestramento fino a quella di utilizzo e dismissione.

È proprio questa estensione temporale e funzionale degli obblighi giuridici a costituire uno degli elementi qualificanti della Convenzione, la quale evita consapevolmente di circoscrivere il proprio ambito di applicazione a specifiche tecnologie o a determinate fasi del processo algoritmico, nella consapevolezza che la rapidità e l'imprevedibilità degli sviluppi tecnologici renderebbero rapidamente obsoleta qualsiasi regolazione eccessivamente puntuale.

In tale prospettiva, il confronto con il regolamento (UE) 2024/1689 sull'intelligenza artificiale – cd. *AI Act* – risulta particolarmente illuminante. Quest'ultimo, fondato sull'art. 114 TFUE e volto primariamente a garantire il corretto funzionamento del mercato unico digitale, adotta un'impostazione marcatamente settoriale e basata sulla classificazione dei sistemi di

1. Cinque firmatari, tra cui almeno tre Stati membri del Consiglio d'Europa.

IA in relazione al livello di rischio. La Convenzione del Consiglio d'Europa, pur non rinunciando a una logica di gestione del rischio, si colloca invece su un piano diverso e complementare, orientato alla tutela dei valori fondamentali e alla preservazione dell'ordine democratico e dello Stato di diritto, indipendentemente dalle ricadute economiche delle tecnologie regolamentate.

Particolarmente significativa è, altresì, la scelta di ancorare la Convenzione non alla creazione di nuovi diritti, bensì alla riaffermazione e all'adattamento di diritti e obblighi già esistenti. In ciò si coglie una precisa opzione di metodo, coerente con la tradizione del Consiglio d'Europa, che evita di frammentare il sistema multilivello di tutela dei diritti fondamentali e mira piuttosto a garantirne la continuità e l'effettività nel mutato contesto tecnologico.

2. Struttura e contenuti essenziali della Convenzione

La «*Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*» del Consiglio d'Europa rappresenta il primo strumento internazionale giuridicamente vincolante che affronti in modo sistemico l'impatto dell'intelligenza artificiale sui valori fondanti dell'ordinamento democratico. Essa si iscrive coerentemente nella tradizione assiologica del Consiglio d'Europa, assumendo quale proprio asse regolativo non la dimensione economica o concorrenziale dell'innovazione tecnologica, bensì la tutela dei diritti umani, della democrazia e dello Stato di diritto, considerati come criteri ordinanti dell'intero spazio tecnologico (Preambolo e art. 1).

Sin dall'art. 1, la Convenzione chiarisce di non voler creare nuovi diritti fondamentali né ridefinire il contenuto di quelli già riconosciuti dagli ordinamenti nazionali e internazionali, ma di mirare piuttosto a garantire che gli obblighi già vigenti in materia di diritti umani trovino piena applicazione anche rispetto alle attività connesse ai sistemi di intelligenza artificiale (art. 1, parr. 1 e 2).

L'IA non è, dunque, configurata come un ambito di eccezione rispetto alla legalità costituzionale e convenzionale, bensì come nuovo spazio di espansione del principio di legalità.

Uno degli elementi più innovativi della Convenzione è l'adozione di una prospettiva estesa all'intero ciclo di vita dei sistemi di intelligenza artificiale, che consente di intercettare i rischi non solo nella fase applicativa, ma anche in quelle di progettazione, sviluppo, addestramento, raccolta dei dati, sperimentazione, distribuzione, utilizzo e dismissione (art. 3). Tale

impostazione risponde all'esigenza di superare una visione statico-funzionale dell'IA, riconoscendone la natura dinamica, adattiva e potenzialmente evolutiva.

Quanto al profilo soggettivo, l'art. 3 delinea un ambito di applicazione ampio, riferito tanto alle autorità pubbliche quanto ai soggetti privati, nella misura in cui le loro attività possano incidere sui diritti umani, sulla democrazia o sullo Stato di diritto (art. 3, par. 1, lett. *a* e *b*). In particolare, gli Stati sono tenuti a esplicitare le modalità attraverso cui intendono dare attuazione agli obblighi convenzionali anche nel settore privato, ferma restando l'impossibilità di eludere o ridurre gli *standard* di tutela stabiliti dal diritto internazionale dei diritti umani. La Convenzione adotta un approccio graduato e differenziato, fondato sulla valutazione della probabilità e della gravità degli impatti negativi dei sistemi di IA sui valori protetti (art. 1, par. 2).

Tale impostazione, lungi dal tradursi in una *deregulation*, attribuisce agli Stati un margine di apprezzamento nell'attuazione delle misure, purché coerente con il nucleo essenziale dei diritti fondamentali.

In questo quadro, assume rilievo centrale il sistema di valutazione e gestione dei rischi previsto dall'art. 16, che impone l'identificazione, l'analisi, la prevenzione e la mitigazione degli impatti sull'intero ciclo di vita dell'IA, anche attraverso strumenti documentati e aggiornabili (art. 16, parr. 1-2).

La Convenzione ammette espressamente, quale *extrema ratio*, il ricorso a moratorie o divieti per quegli usi dei sistemi di intelligenza artificiale ritenuti incompatibili con il rispetto dei diritti umani, della democrazia o dello Stato di diritto (art. 16, par. 4).

Il cuore assiologico della Convenzione è rappresentato dalla riaffermazione della centralità della persona, declinata attraverso il principio di dignità umana e autonomia individuale (art. 7). I sistemi di IA non devono condurre a processi di de-umanizzazione decisionale, né ridurre l'individuo a mero oggetto di classificazione algoritmica, né interferire con la capacità di autodeterminazione.

A tale principio si affianca il divieto di discriminazioni algoritmiche e l'obbligo di promuovere l'uguaglianza sostanziale, anche attraverso l'adozione di misure idonee a contrastare *bias* strutturali e sistemici incorporati nei dati o nei modelli decisionali (artt. 10 e 17). La tutela è rafforzata per i soggetti particolarmente vulnerabili, in primo luogo i minori e le persone con disabilità, rispetto ai quali gli Stati sono tenuti a considerare bisogni specifici e rischi accresciuti (art. 18).

La Convenzione valorizza in modo significativo il principio di trasparenza e supervisione, imponendo che i sistemi di IA siano governati in modo comprensibile, spiegabile e interpretabile, nei limiti

della fattibilità tecnica e nel rispetto degli interessi concorrenti costituzionalmente protetti (art. 8). La trasparenza non costituisce un valore isolato, ma una condizione funzionale all'effettività delle tutele giurisdizionali.

In stretta connessione con tale principio, l'art. 9 consacra il principio di responsabilità e *accountability*, imponendo agli Stati di assicurare l'esistenza di meccanismi idonei a individuare e imputare le decisioni e gli effetti prodotti dai sistemi di IA a soggetti identificabili (art. 9). Ne discende il rafforzamento del diritto a rimedi effettivi in caso di violazioni dei diritti umani, nonché l'esigenza di preservare adeguate garanzie procedurali, incluse forme di controllo umano significativo sulle decisioni automatizzate (artt. 14 e 15).

Particolare rilievo assume la tutela dell'integrità dei processi democratici e dello Stato di diritto. La Convenzione evidenzia i rischi connessi all'uso distorto dell'IA in ambiti quali la formazione dell'opinione pubblica, il pluralismo informativo, i processi elettorali e l'indipendenza della funzione giudiziaria (art. 5). La protezione della capacità degli individui di formare liberamente le proprie opinioni emerge quale presupposto imprescindibile della legittimità democratica nell'era algoritmica.

Sul piano istituzionale, la Convenzione istituisce un articolato meccanismo di monitoraggio e cooperazione internazionale, fondato sulla Conferenza delle Parti (art. 23), sugli obblighi di rendicontazione (art. 24), sulla cooperazione tra Stati (art. 25) e sulla previsione di autorità di supervisione indipendenti ed effettive (art. 26).

3. Governance dell'intelligenza artificiale e centralità della persona

Nel suo complesso, la *Framework Convention on Artificial Intelligence* realizza una sintesi avanzata tra innovazione tecnologica e continuità dei valori costituzionali, riaffermando un principio di fondo: anche nell'era degli algoritmi, la tecnica resta soggetta al diritto e il diritto resta al servizio della persona. Essa costituisce, pertanto, un punto di riferimento imprescindibile per l'evoluzione della *governance* democratica dell'intelligenza artificiale, tanto a livello europeo quanto globale.

Il nodo dell'ambito applicativo della Convenzione assume rilievo non soltanto sotto il profilo tecnico-giuridico, ma anche quale indicatore della visione complessiva del rapporto tra autorità pubbliche, soggetti privati e tecnologia. La scelta di concentrare l'obbligo convenzionale in modo prioritario sull'uso dei sistemi di intelligenza artificiale da parte delle autorità pubbliche – nonché dei soggetti privati che

agiscono per loro conto – evidenzia l'attenzione del Consiglio d'Europa per le aree nelle quali il potere tecnologico si intreccia in modo più diretto con l'esercizio della sovranità e con l'incidenza sulle sfere giuridiche individuali. In tali ambiti, la mediazione algoritmica non rappresenta un semplice strumento tecnico, ma si configura come fattore potenzialmente idoneo a ridisegnare le modalità di esercizio del potere pubblico, con conseguenze dirette sulla legalità dell'azione amministrativa, sulla garanzia del giusto processo e sull'effettività della tutela giurisdizionale.

Il margine lasciato agli Stati nella regolazione delle attività svolte da attori privati non equivale, tuttavia, a una deresponsabilizzazione. Al contrario, esso si inserisce in una concezione sussidiaria e multilivello della protezione dei diritti fondamentali, nella quale permane fermo l'obbligo degli Stati di assicurare che nessuna attività, ancorché formalmente privata, si traduca in una violazione degli *standard* internazionali di tutela dei diritti umani. La discrezionalità riconosciuta agli Stati nell'attuazione delle disposizioni convenzionali non attenua, dunque, la natura vincolante degli obblighi di risultato imposti dalla Convenzione, ma richiede un'attenta calibrazione degli strumenti normativi, amministrativi e regolatori interni in relazione ai diversi contesti di utilizzo dell'intelligenza artificiale.

Particolarmente significativa, in questo quadro, è anche la delimitazione delle esclusioni dall'ambito applicativo, in specie con riferimento alla sicurezza e alla difesa nazionale. Tale esclusione, che ricalca un limite tradizionale dell'azione del Consiglio d'Europa, non può essere letta come una zona franca sottratta al diritto internazionale dei diritti umani. La salvaguardia degli interessi di sicurezza nazionale resta infatti soggetta ai principi di legalità, necessità e proporzionalità, che costituiscono parte integrante del patrimonio convenzionale e che si applicano anche in ambito tecnologico, specie laddove l'uso dell'intelligenza artificiale possa incidere in modo significativo sulla vita privata, sulla libertà di espressione o sulla libertà personale.

Il centro di gravità della Convenzione è rappresentato dagli obblighi generali e dai principi fondamentali che devono informare l'intero ciclo di vita dei sistemi di IA. In questo ambito, emerge con particolare chiarezza la scelta di collocare la dignità umana e l'autonomia individuale quali valori assiologici primari, rispetto ai quali la tecnologia deve mantenere una funzione meramente strumentale. L'autonomia dell'individuo, intesa come capacità di autodeterminarsi e di compiere scelte libere e consapevoli, rischia infatti di essere compromessa da sistemi algoritmici opachi, manipolativi o eccessivamente pervasivi, soprattutto quando essi operano in settori sensibili

come l'accesso ai servizi pubblici, la sorveglianza, la giustizia predittiva o la sanità.

La valorizzazione del principio di trasparenza e del controllo umano costituisce, in tale prospettiva, una risposta diretta al rischio di una delega incontrollata alla macchina. La Convenzione non si limita a richiedere la mera conoscibilità dell'esistenza di un sistema di intelligenza artificiale, ma esige che le modalità di funzionamento e i criteri decisionali siano intelligibili in misura tale da consentire un controllo effettivo, sia da parte delle autorità competenti sia da parte degli individui interessati. La trasparenza, così intesa, assume una funzione eminentemente garantista, poiché si pone come presupposto necessario per l'esercizio del diritto di difesa, per l'accesso a rimedi effettivi e per la stessa legittimazione democratica delle decisioni automatizzate.

In stretta connessione con la trasparenza si colloca il principio di *accountability*, che impone l'individuazione chiara dei soggetti responsabili lungo l'intero ciclo di vita del sistema di intelligenza artificiale. La complessità tecnica degli algoritmi, infatti, non può tradursi in una diluizione della responsabilità giuridica. La Convenzione afferma con nettezza che l'attribuzione delle decisioni al sistema non esonera dal dovere di rendere conto degli effetti prodotti, riaffermando così un principio cardine dello Stato di diritto: ogni esercizio di potere che incida sui diritti deve essere giuridicamente imputabile.

Ampio spazio è dedicato alla prevenzione delle discriminazioni, incluse quelle indirette o sistemiche che possono derivare dall'uso di dati distorti o da modelli di addestramento non rappresentativi. In questo ambito, la Convenzione si colloca in continuità con il diritto internazionale e regionale antidiscriminatorio, ma ne estende la portata al contesto algoritmico, riconoscendo che l'apparente neutralità della decisione automatizzata può mascherare meccanismi di esclusione particolarmente insidiosi. Il riferimento espresso alla parità di genere e alla tutela dei soggetti in condizione di vulnerabilità rafforza l'impostazione sostanziale del testo, che mira a contrastare non solo le discriminazioni manifeste, ma anche quelle strutturali e latenti.

Non meno rilevante è la disciplina dei rimedi e delle garanzie procedurali, che costituisce una delle sezioni più delicate e innovative della Convenzione. In un contesto caratterizzato da una marcata asimmetria informativa tra sviluppatori, utilizzatori e destinatari dei sistemi di IA, l'accesso a informazioni adeguate e la possibilità di contestare le decisioni automatizzate diventano condizioni essenziali per la tutela effettiva dei diritti. La Convenzione non istituisce nuovi meccanismi giurisdizionali sovranazionali, ma impone agli Stati di assicurare che i rimedi esistenti siano pienamente operativi anche nei casi in cui la

violazione derivi dall'uso dell'intelligenza artificiale, riaffermando il principio di continuità della protezione dei diritti umani.

Nel solco della prevenzione si colloca anche l'approccio alla gestione del rischio, che permea l'intera architettura convenzionale. La scelta di non definire rigidamente il concetto di "rischio", lasciando agli Stati la possibilità di adottare metodologie e strumenti differenziati, esprime la consapevolezza della varietà dei contesti applicativi e della rapidità dell'evoluzione tecnologica. Tuttavia, tale flessibilità non si traduce in indeterminatezza: gli Stati sono chiamati a valutare in modo sistematico e documentato i potenziali impatti dei sistemi di IA sui diritti umani, sulla democrazia e sullo Stato di diritto, adottando misure proporzionate alla gravità e alla probabilità dei rischi individuati.

Il meccanismo di *follow-up* affidato alla Conferenza delle Parti rappresenta un ulteriore elemento di equilibrio tra vincolatività e cooperazione. Esso non si configura come un organo giurisdizionale, ma come una sede di monitoraggio, dialogo e scambio di buone pratiche, in linea con l'approccio tradizionale del Consiglio d'Europa. La funzione consultiva e interpretativa della Conferenza è destinata a svolgere un ruolo significativo nell'orientare l'attuazione della Convenzione e nell'assicurare una lettura evolutiva delle sue disposizioni, capace di adattarsi ai mutamenti tecnologici senza tradirne l'impianto valoriale.

Particolare rilievo sistemico assume, infine, il coinvolgimento dell'Unione europea, che ha partecipato attivamente ai negoziati e ha sottoscritto la Convenzione. Le disposizioni dedicate ai rapporti tra diritto dell'Unione e Convenzione riflettono un delicato equilibrio tra l'autonomia dell'ordinamento europeo e l'esigenza di garantire la coerenza degli *standard* di tutela dei diritti fondamentali. In tale contesto, la Convenzione si pone come strumento potenzialmente idoneo a rafforzare il dialogo tra la Corte di giustizia dell'UE e la Corte europea dei diritti dell'uomo, anche alla luce delle questioni emergenti legate all'intelligenza artificiale.

Nel suo complesso, la Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale si colloca come un tassello essenziale nel processo di adattamento del diritto ai mutamenti tecnologici. Il suo contributo più significativo non risiede nella creazione di nuove categorie giuridiche, ma nella capacità di riaffermare, con rinnovata forza, la centralità della persona e la funzione garantista del diritto in un contesto in cui la tecnologia rischia di sottrarsi al controllo democratico. In tal senso, essa rappresenta un passo decisivo verso una regolazione dell'IA autenticamente umano-centrica, nella quale l'innovazione è chiamata a porsi al servizio dei diritti, della democrazia e dello Stato di diritto, e non viceversa.

4. Discrezionalità algoritmica, depoliticizzazione del potere decisionale e ruolo della giurisdizione

È proprio nel rapporto tra intelligenza artificiale e discrezionalità che la Convenzione rivela, se letta in chiave critico-politica, una delle tensioni più profonde dell'impianto regolatorio. L'introduzione di sistemi algoritmici nei processi decisionali pubblici non elimina la discrezionalità, ma ne trasforma radicalmente la natura e i luoghi di esercizio. La discrezionalità "migra" a monte, nelle fasi di progettazione, selezione dei dati, definizione delle finalità e delle metriche di ottimizzazione, tutte attività che il *Rapporto* esplicativo della Convenzione include pienamente nel ciclo di vita dei sistemi di IA².

In tal modo, scelte intrinsecamente politiche e di valore tendono a presentarsi come necessità tecniche, sottraendosi al conflitto democratico e al controllo pubblico. Il rischio che si profila non è semplicemente quello di decisioni errate o discriminatorie, ma quello di una discrezionalità mimetizzata, resa opaca dal linguaggio dell'efficienza e dell'oggettività.

In questa prospettiva, la trasparenza e l'intelligibilità, pur centrali nell'impianto della Convenzione, non possono essere lette come meri requisiti informativi. Esse costituiscono, piuttosto, presidi contro una depoliticizzazione del potere che rischia di spostare il baricentro delle decisioni verso spazi tecnici scarsamente sindacabili, talora controllati da attori privati. La *governance* per *standard*, documentazione e *audit*, richiamata dal *Rapporto* esplicativo come strumento di affidabilità, può trasformarsi in una forma di managerialismo dei diritti, in cui la correttezza procedurale finisce per sostituire la giustizia sostanziale degli esiti.

Per la giurisdizione, la sfida è particolarmente rilevante. Il giudice non è chiamato né a ratificare l'*output* algoritmico né a respingerlo in quanto tale, ma a ricondurre a responsabilità giuridica quella discrezionalità che l'algoritmo tende a occultare. La discrezionalità algoritmica deve essere trattata come discrezionalità in senso pieno: oggetto di motivazione, di controllo e di valutazione di compatibilità con i diritti fondamentali. In questo senso, il *Rapporto* esplicativo alla Convenzione sembra affidare alla giurisdizione un compito eminentemente politico-costituzionale: impedire che l'automazione decisionale si

traduca in una sottrazione dell'esercizio del potere ai presidi della legalità costituzionale, della responsabilità giuridica e del controllo democratico.

Particolare attenzione merita, in questo quadro, l'impatto dell'IA sull'esercizio della funzione giurisdizionale e, più in generale, sull'amministrazione della giustizia. Se è vero che la Convenzione quadro non si presenta come uno strumento settoriale, ma come una disciplina trasversale degli effetti dell'IA sui valori fondanti dell'ordinamento democratico, è altrettanto vero che il settore giustizia rappresenta uno dei luoghi nei quali tali effetti si manifestano con particolare intensità e delicatezza. L'uso di sistemi algoritmici in ambito giudiziario, infatti, incide direttamente su diritti fondamentali di primario rilievo, sul principio di indipendenza del giudice, sulle garanzie del giusto processo e, in ultima analisi, sulla stessa fiducia dei cittadini nell'amministrazione della giustizia.

Non è un caso che il Consiglio d'Europa, ben prima dell'adozione della Convenzione quadro, abbia sviluppato una riflessione specifica sull'uso dell'intelligenza artificiale nei sistemi giudiziari attraverso il lavoro della Commissione europea per l'efficienza della giustizia (CEPEJ). Già con l'adozione della «*Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*» del 2018³, il Consiglio d'Europa ha individuato principi quali il rispetto dei diritti fondamentali, la non discriminazione, la qualità e la sicurezza dei dati, la trasparenza e il controllo umano come condizioni imprescindibili per qualsiasi utilizzo dell'IA nella giurisdizione. Successivamente, la CEPEJ ha ulteriormente consolidato tale indirizzo attraverso la creazione di un Consiglio consultivo sull'IA e l'elaborazione di documenti specificamente rivolti ai professionisti della giustizia, come la nota informativa sull'«*Uso dell'intelligenza artificiale generativa da parte degli operatori giudiziari in ambito lavorativo*»⁴.

Queste iniziative trovano oggi un solido fondamento giuridico nella Convenzione quadro, che, pur senza disciplinare in modo puntuale le singole applicazioni giudiziarie dell'IA, fornisce un quadro normativo di riferimento idoneo a orientarne l'utilizzo in modo compatibile con lo Stato di diritto. In particolare, gli obblighi generali di tutela dei diritti umani (artt. 4 e 5), il principio di rispetto della dignità umana e dell'autonomia individuale (art. 7), le garanzie di trasparenza e sorveglianza (art. 8), nonché l'obbligo

2. *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Vilnius, 5 settembre 2024 (<https://rm.coe.int/1680afae67>).

3. CEPEJ, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Strasburgo, 3-4 dicembre 2018 (<https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>).

4. *Use of Generative Artificial Intelligence (AI) by judicial professionals in a work-related context*, Strasburgo, 12 febbraio 2024 (<https://rm.coe.int/cepej-gt-cyberjust-2023-5final-en-note-on-generative-ai/1680ae8e01>).

di assicurare mezzi di ricorso effettivi e garanzie procedurali (artt. 14 e 15) assumono un rilievo centrale quando l'IA è impiegata a supporto dell'attività giurisdizionale o in fasi procedurali che incidono direttamente sulla posizione giuridica delle parti.

In tale prospettiva, la Convenzione rafforza un principio che può essere considerato immanente al costituzionalismo europeo: nessun sistema di intelligenza artificiale può sostituire il giudice nell'esercizio della funzione decisionale, né può essere utilizzato in modo tale da svuotare di contenuto l'obbligo di motivazione, il contraddittorio o il diritto delle parti a comprendere e contestare le decisioni che le riguardano. L'esigenza di rendere conoscibili le logiche di funzionamento dei sistemi di IA, richiamata espressamente dall'art. 14, par. 2, lett. b della Convenzione, assume in ambito giudiziario una valenza particolarmente incisiva, poiché si traduce nel diritto dell'interessato di comprendere se e in che misura una decisione sia stata influenzata da un sistema algoritmico e di contestarne efficacemente l'utilizzo o i risultati.

Il raccordo tra intelligenza artificiale e funzione giurisdizionale non può, inoltre, prescindere dalla giurisprudenza della Corte europea dei diritti dell'uomo in materia di equo processo, tutela della vita privata e limiti all'automazione decisionale. Sebbene la Corte non si sia ancora pronunciata in modo diretto e sistematico sull'uso giudiziario dei sistemi di IA, le sue sentenze in tema di sorveglianza di massa, raccolta e trattamento automatizzato dei dati e sicurezza nazionale delineano principi destinati a riverberarsi inevitabilmente anche sul terreno della giustizia algoritmica. In tali pronunce, la Corte ha infatti ribadito che l'uso di tecnologie avanzate non può mai compromettere l'essenza dei diritti garantiti dalla Convenzione europea dei diritti dell'uomo, né sottrarre l'azione pubblica a un controllo effettivo, indipendente e giurisdizionale⁵.

In tale prospettiva, la Convenzione quadro non si limita a imporre obblighi vincolanti al legislatore

e all'amministrazione, ma fornisce altresì alle giurisdizioni nazionali una griglia interpretativa di riferimento per valutare la compatibilità delle applicazioni algoritmiche con i principi del giusto processo e dell'indipendenza giudiziaria. Essa rafforza il ruolo del giudice quale garante ultimo dei diritti fondamentali anche nello spazio digitale e riafferma che la tecnologia, lungi dal costituire una fonte autonoma di legittimazione della decisione, deve restare uno strumento ancillare, sottoposto al controllo umano e alla responsabilità istituzionale.

Alla luce di tali considerazioni, il contributo della Convenzione quadro alla regolazione dell'intelligenza artificiale non può essere compreso appieno senza tenere conto della sua ricaduta sulla funzione giurisdizionale. Proprio nel settore della giustizia si misura, forse più che altrove, la capacità dell'ordinamento di governare l'innovazione tecnologica senza sacrificare i presidi essenziali dello Stato di diritto. La scelta del Consiglio d'Europa di inscrivere l'uso dell'IA entro il perimetro dei diritti, delle garanzie procedurali e del controllo giurisdizionale rappresenta, in questa prospettiva, un passaggio decisivo per evitare che l'automazione si traduca in una forma inedita di opacità del potere, sottratta al vaglio del diritto.

Il capitolo quarto della Convenzione, dedicato ai rimedi, consolida questa impostazione. Quando una decisione incide in modo significativo sui diritti fondamentali ed è sostanzialmente orientata dall'uso dell'IA, devono essere garantiti rimedi effettivi, accessibili e comprensibili. La trasparenza diviene così condizione essenziale del diritto di difesa: senza conoscenza del ruolo svolto dal sistema algoritmico, il diritto al ricorso e alla contestazione risulta svuotato.

Il controllo umano – *ex ante* o *ex post* – assume in questa cornice un significato che va ben oltre la dimensione tecnica, intrecciandosi con i principi del giusto processo, della motivazione e dell'attribuibilità delle decisioni.

5. Corte Edu, *Big Brother Watch et al. c. Regno Unito* [GC], n. 58170/13, 62322/14, 24960/15, 25 maggio 2021; *Centrum för Rättvisa c. Svezia* [GC], n. 35252/08.