
Intelligenza artificiale: un trattato internazionale basato sull'"approccio umano"

Sta per entrare in vigore la “Convenzione quadro del Consiglio d’Europa sull’intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto”, un testo che fa storia perché è il primo trattato internazionale giuridicamente vincolante che nasce per garantire il rispetto delle norme giuridiche in materia di diritti umani, democrazia e Stato di diritto nell’utilizzo dei sistemi di intelligenza artificiale. A redigere la Convenzione, in due anni di lavoro, è stato il Comitato sull’intelligenza artificiale, organismo intergovernativo che ha riunito i delegati di 46 Stati membri del Consiglio d’Europa e di 11 Stati osservatori (Argentina, Australia, Canada, Costa Rica, Giappone, Israele, Messico, Perù, Santa Sede, Stati Uniti d’America e Uruguay), nonché l’Unione europea e oltre 70 osservatori del settore privato, della società civile e del mondo accademico. Adottata lo scorso maggio a Strasburgo durante l’incontro ministeriale annuale del Comitato dei ministri del Consiglio d’Europa (cioè i ministri degli esteri dei 46 Stati membri dell’Organizzazione), in occasione dell’incontro dei ministri della giustizia a Vilnius, oggi, giovedì 5 settembre, sarà ufficialmente aperta alla firma (per l’effettiva entrata in vigore occorrono almeno 5 ratifiche, di cui almeno tre ad opera di Stati membri del Consiglio). Potranno aderire alla convenzione anche Paesi non europei. L’obiettivo di questo testo è garantire che le attività svolte nell’ambito del ciclo di vita dei sistemi di intelligenza artificiale siano pienamente coerenti con i diritti umani, la democrazia e lo Stato di diritto, pur favorendo il progresso tecnologico e l’innovazione. Sono oggetto della convenzione l’uso dei sistemi di IA nel settore pubblico, comprese le aziende che operano per suo conto, e nel settore privato. I Paesi che vi aderiscono sono vincolati quindi ad adottare misure per identificare, valutare, prevenire e attenuare i rischi dell’uso dei sistemi di IA e a stabilire quali misure vengano messe in campo quando i sistemi presentano rischi incompatibili con le normative internazionali in materia di diritti umani, democrazia e Stato di diritto. Il testo, che per ora è disponibile sul sito del Consiglio solo in inglese e in francese, (https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680afb11f) è composto da 36 articoli, raggruppati in 8 capitoli principali (Disposizioni generali, Obblighi generali, Principi relativi alle attività del ciclo di vita dei sistemi di intelligenza artificiale, Rimedi, Valutazione e mitigazione dei rischi e degli impatti negativi, Attuazione della Convenzione. Meccanismo di verifica e cooperazione, Clausole finali). L’elemento più significativo è il fatto che la pietra angolare della norma diventa l’approccio “umano-centrico”, dal momento che, all’articolo 7, si dice che chi aderisce alla convenzione si impegna ad “adottare o mantenere misure volte a rispettare la dignità umana e l’autonomia individuale in relazione alle attività svolte nell’ambito del ciclo di vita dei sistemi di intelligenza artificiale”. Accanto a questo, principi cardine posti dalla convenzione sono la trasparenza e il controllo - per cui processi e funzionamento dei sistemi IA dovrebbero essere comprensibili e accessibili -, il principio di responsabilità, di uguaglianza e di non discriminazione, il diritto alla privacy e alla protezione dei dati personali, l’affidabilità e la sicurezza dell’innovazione. Inoltre la Convenzione vincola i Paesi aderenti a definire garanzie procedurali per le vittime di violazioni dei diritti umani legate all’utilizzo di sistemi di IA. In relazione ai rischi per la democrazia, il trattato richiede alle parti di adottare misure per assicurare che i sistemi di IA non siano utilizzati per compromettere le istituzioni e i processi democratici, tra cui il principio di separazione dei poteri, il rispetto dell’indipendenza giudiziaria e l’accesso alla giustizia. Escluse dalla sfera di applicazione sono le attività degli Stati aderenti riguardo la protezione degli interessi di sicurezza nazionale, la difesa nazionale, le attività di ricerca e sviluppo dei sistemi stessi, salvo nei casi in cui interferiscano con i diritti umani, la democrazia e lo Stato di diritto. Una conferenza delle parti avrà il compito di monitorare l’attuazione della Convenzione, ma ad ogni singolo aderente è richiesto di istituire un meccanismo di controllo indipendente per controllare la conformità alla convenzione, sensibilizzare l’opinione pubblica, stimolare il dibattito e il confronto sugli usi della tecnologia IA. “Negoziare uno strumento vincolante in un processo così inclusivo e in una scadenza

molto breve è stato intenso e talvolta impegnativo”, ha raccontato lo svizzero **Thomas Schneider**, che ha presieduto i lavori del comitato. “Abbiamo dovuto trovare modi per colmare le differenze tra i sistemi e le tradizioni giuridiche degli Stati, e anche tra le interpretazioni di alcuni diritti umani. Abbiamo anche dovuto gestire una serie di aspettative su come sviluppare uno strumento legale che avrebbe dovuto avere un impatto sulla governance dell'IA a livello globale”. Il prossimo passo sarà l'elaborazione di una metodologia per la valutazione del rischio e dell'impatto dei sistemi di intelligenza artificiale sui diritti umani, sulla democrazia e sullo stato di diritto. I critici della convenzione hanno segnalato come nel corso delle negoziazioni, si sia indebolito il testo su due aspetti: il primo è nell'aver sottratto all'ambito di competenza della Convenzione tutto il tema della sicurezza nazionale; il secondo riguarda il doppio standard per attori pubblici e attori privati, per cui i primi sono vincolati, mentre rispetto al settore privato si lascia alle parti il compito di decidere come far applicare i principi e gli obblighi alle attività degli attori privati. Si sarebbe trattato di compromessi necessari al raggiungimento dell'obiettivo, cioè la definizione del primo regolamento internazionale sull'IA, applicato anche al di fuori delle frontiere dell'Ue. Per la segretaria generale del Consiglio d'Europa, **Marija Pejčinović**, la Convenzione è “una risposta alla necessità di disporre di una norma di diritto internazionale sostenuta da Stati di diversi continenti uniti da valori comuni, che consenta di trarre vantaggio dall'intelligenza artificiale, riducendo al contempo i rischi che questa presenta”. Che si tratti di “una risposta”, lo conferma il fatto che lo scorso dicembre, anche l'Unione europea ha varato il suo “Ai Act”. In realtà anche l'Onu si sta muovendo: il 22-23 settembre a New York si terrà il “Summit del futuro”, convocato dal segretario generale Antonio Guterres per portare i leader mondiali a riconcentrarsi sul raggiungimento degli obiettivi di sviluppo del millennio. Al summit, i leader mondiali adotteranno il Patto per il futuro, che includerà un Global Digital Compact e una Dichiarazione sulle generazioni future (tra i membri dell'advisory board nominato dall'Onu per formulare raccomandazioni per il Patto globale sul digitale e sul tema dell'intelligenza artificiale, c'è padre Paolo Benanti, francescano, professore di bioetica ed etica delle tecnologie all'Università gregoriana). Ancora non è pubblico il testo definitivo che arriverà sul tavolo del summit: sono disponibili le bozze e sono emerse le difficoltà negoziali tra approcci polarmente opposti come quello di Cina e Usa. C'è anche l'Africa che si sta mettendo in moto per sviluppare una propria cornice normativa che “rifletta il suo contesto unico, per garantire che lo sviluppo e l'uso dell'IA sia responsabile, etico e benefico per tutti”, si leggeva su una testata nigeriana indipendente, ConsumerConnect, in un editoriale apparso a fine agosto. Le regole sono più che necessarie, ardua l'armonizzazione degli approcci, a portata di mano la logica del minimo comune denominatore, molto rischioso per un ambito così delicato come quello dell'intelligenza artificiale e le sue applicazioni.

Sarah Numico