NOMOS
Le attualità nel diritto
Quadrimestrale di Teoria generale,
Diritto pubblico comparato e Storia costituzionale

FLAVIA ZORZI GIUSTINIANI*

LA CONVENZIONE QUADRO SULL'INTELLIGENZA ARTIFICIALE E I DIRITTI UMANI, LA DEMOCRAZIA E LO STATO DI DIRITTO E L'*INTERNATIONAL CYBERSPACE AND DIGITAL POLICY STRATEGY* STATUNITENSE **

Il **17 maggio** scorso, durante la sua centotrentatreesima sessione, il Comitato dei Ministri del Consiglio d'Europa ha adottato all'unanimità la Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto[1]. La Convenzione, finalizzata al termine di due anni di intensi negoziati portati avanti dal Comitato sull'intelligenza artificiale, costituisce il primo strumento internazionale a carattere pattizio finalizzato a regolare l'uso dell'intelligenza artificiale (IA) nel rispetto dei diritti fondamentali e della *rule of law*.

La Convenzione crea un quadro giuridico comune che intende affrontare le varie questioni etiche, giuridiche e sociali derivanti dall'utilizzo dei sistemi di IA garantendo che tutte le attività svolte nell'ambito del ciclo di vita dei suddetti sistemi «are fully consistent with human rights, democracy, and the rule of law»[2]. Si basa su regole e principi generali quali la trasparenza, la non discriminazione e la tutela della privacy.

I sistemi di IA non si basano sulla corrispondente definizione, assai letterale, contenuta nell'omonimo regolamento dell'Unione europea, bensì su quella adottata da ultimo dall'OCSE nel novembre 2023[3]. Benché le due definizioni siano sostanzialmente equivalenti, basandosi entrambe sugli aspetti chiave dei sistemi di IA (autonomia e adattabilità variabili, capacità di inferenza e generazione di previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali), la

* Professoressa associata di Diritto dell'Unione europea – Università degli Studi Link, Roma.

** Contributo sottoposto a *peer review*.

1 Cfr. CM(2024)52-final. La Convenzione è stata poi aperta alla firma il 5 settembre scorso a Vilnius (Lituania), in occasione di una conferenza dei ministri della Giustizia con la previsione di una duplice modalità di attuazione. Gli Stati firmatari possono infatti decidere se aderire alle disposizioni pattizie come risultano dal trattato oppure impegnarsi ad attuare altre misure paragonabili a quelle del documento, purché tutelino "i diritti umani, la democrazia e lo Stato di diritto". Siffatta duplice modalità è stata proposta per tenere conto delle diversità sussistente negli ordinamenti giuridici dei Paesi firmatari.

2 The AI Convention (15 March 2024), page 4, Article 1 (Object and purpose).

3 Cfr. OCSE, *Recommendation of the Council on Artificial Intelligence*, I, OECD/LEGAL/0449, 22 maggio 2019, emendata da ultimo il 3 maggio 2024.

ISSN 2279-7238

preferenza per la nozione dell'OCSE si deve all'intento di potenziare la cooperazione internazionale in materia di IA e di facilitare gli sforzi per armonizzarne la *governance* a livello globale.

Il testo, come emerge dal suo stesso preambolo, nasce dalla consapevolezza degli effetti che gli sviluppi scientifici e tecnologici, e in particolare l'utilizzo delle tecnologie definite o comunque rientranti nella nozione di IA, possono avere per accrescere la prosperità del genere umano così come quella dei singoli individui, spingendo l'innovazione e aumentando la partecipazione su vasta scala dei relativi progressi economici, sociali e culturali. Nel contempo, tuttavia, il testo mette l'accento anche sugli effetti negativi e repressivi per i singoli che un cattivo uso dell'IA può determinare, con il rischio di limitazione del libero arbitrio come pure di ricorso generalizzato a forme illegittime di sorveglianza e di censura. L'approccio della Convenzione è pertanto sostanzialmente umano-centrico, giacché il testo si concentra in via esclusiva sugli aspetti attinenti alla promozione e alla protezione dei diritti umani, tralasciando invece del tutto la regolazione degli aspetti economici o di mercato dei sistemi di IA. In altri termini, la Convenzione non mira a regolamentare tutte le attività all'interno del ciclo di vita dei sistemi di IA ma solo quelle che possono interferire con i diritti umani, la democrazia e lo Stato di diritto. La portata oggettiva non coincide quindi, a differenza del regolamento dell'Unione europea sull'intelligenza artificiale, con specifici modelli, sistemi o pratiche di IA, ma con le singole attività all'interno del ciclo di vita dell'IA e l'impatto che queste possono avere anche indipendentemente dal rischio che l'intero sistema presenta.

Il processo negoziale, non a caso, si è contraddistinto per la molteplicità degli attori coinvolti, tra i quali si ricordano, oltre ai 46 Stati membri del Consiglio d'Europa, undici Paesi non membri (quali Argentina, Australia, Canada, Costa Rica, Giappone, Israele, Messico, Perù, Santa Sede, Stati Uniti d'America e Uruguay)[4], nonché l'Unione europea e svariati rappresentanti del settore privato, della società civile e del mondo accademico, che vi hanno preso parte in qualità di osservatori. Ciò si deve all'intento del Comitato dei Ministri di fissare standard globali e di incoraggiare pertanto anche i Paesi non membri del Consiglio d'Europa a sottoscrivere il costituendo trattato, come è poi effettivamente avvenuto ad esempio con gli Stati Uniti e l'Unione europea.

Una siffatta apertura ha tuttavia comportato anche una minore cogenza dell'articolato. Così è avvenuto in primis con riferimento all'ambito di applicazione *ratione personae* e *ratione materiae* della Convenzione. Sotto il primo profilo, la Convenzione si applica principalmente alle pubbliche autorità e agli attori privati che agiscono per conto delle prime, rimettendo alla discrezionalità delle singole parti firmatarie il trattamento da riservare ai privati[5]. Le parti

[4] Secondo il *Draft Explanatory Report* dell'accordo il Comitato dei Ministri ha deciso infatti "*to allow for the inclusion in the negotiations of the European Union and interested non-European States sharing the values and aims of the Council of Europe – States from around the globe, namely Argentina, Australia, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru, the United States of America and Uruguay, joined the process of negotiations in the CAI and participated in the elaboration of this Framework Convention as observer States*" (disponibile al link https://rm.coe.int/-1497-10-1b-committee-on-artificial-intelligence-cai-b-draft-framework-convention/1680af0734&format=native / https://rm.coe.int/-1497-10-1b-comite-sur-l-intelligence-artificielle-cai-b-projet-de-convention/1680af0733&format=native).

[5] Cfr. art. 3 par. 1 A e B della Convenzione.

contraenti si impegnano infatti ad affrontare i rischi e gli impatti derivanti dalle attività svolte da attori privati nell'ambito del ciclo di vita dei sistemi di IA in modo conforme all'oggetto e allo scopo della Convenzione, ma hanno la possibilità di scegliere se applicare gli obblighi della Convenzione o adottare altre misure appropriate.

Ratione materiae, invece, lo strumento non è applicabile a questioni concernenti la difesa nazionale[6], né alle attività di ricerca e sviluppo relative ai sistemi di IA che non sono stati ancora resi disponibili all'utilizzo, «unless testing or similar activities are undertaken in such a way that they have the potential to interfere with human rights, democracy and the rule of law»[7].

Sul rispetto degli obblighi convenzionali dovranno poi vigilare, a livello nazionale, appositi meccanismi di controllo efficaci designati dalle singole parti contraenti e dotati delle competenze e delle risorse necessarie per svolgere il proprio compito in modo indipendente ed imparziale[8].

Come l'omonimo regolamento europeo, anche la Convenzione del Consiglio d'Europa è improntata ad un approccio *risk-based* alla regolazione dei sistemi di IA e include specifiche disposizioni per le valutazioni del rischio e d'impatto come pure misure per la mitigazione del rischio.

Dopo una prima parte dedicata alle disposizioni generali e agli obblighi delle parti, la Convenzione fissa una serie di principi, che dovranno essere attuati a livello nazionale e che si caratterizzano per l'elevata generalità nell'ottica di consentirne una migliore e agevole applicazione in contesti soggetti a rapidi cambiamenti. Il primo principio richiede misure che rispettino la dignità umana e l'autonomia individuale. In particolare, l'uso di sistemi di IA non dovrebbe portare alla disumanizzazione degli individui, minando la loro capacità di agire in modo autonomo o riducendoli a meri punti dati. Inoltre, i sistemi di IA non dovrebbero essere antropomorfizzati in modo da interferire con la dignità umana. Il secondo principio attiene invece alla trasparenza e alla supervisione dei sistemi di IA. Al riguardo la Convenzione richiede l'adozione o il mantenimento di misure per garantire la trasparenza e il monitoraggio adattato a contesti e rischi specifici, inclusa l'identificazione dei contenuti generati dall'intelligenza artificiale. Segue il principio di *accountability* e responsabilità, che richiede l'istituzione di appositi meccanismi atti a valutare le responsabilità dei vari enti come dei singoli coinvolti nel ciclo di vita dei sistemi di IA per eventuali impatti negativi sui diritti umani, sulla democrazia o sullo stato di diritto. Ancora, vengono enunciati i principi di uguaglianza e non discriminazione, *privacy* e protezione dei dati personali, affidabilità (basata su standard tecnici e misure in termini di robustezza, accuratezza, integrità dei dati e sicurezza informatica), nonché innovazione sicura (volta a consentire, ove opportuno, la creazione di ambienti controllati, quali ad esempio *sandbox* regolamentari, per lo sviluppo, la sperimentazione e il test dei sistemi di intelligenza artificiale sotto la supervisione delle autorità competenti).

[6] Cfr. art. 3, par. 4 della Convenzione.
[7] Cfr. art. 3, par. 3 della Convenzione.
[8] Cfr. art. 26 della Convenzione.

In conclusione, l'adozione della Convenzione in discorso costituisce sicuramente uno sviluppo importante nel settore, con ricadute potenzialmente globali come globale è potenzialmente anche la sua portata. Nondimeno, anche a causa della genericità dei suoi obblighi e dell'ampio margine di discrezionalità che è lasciato alle parti contraenti, il suo impatto rischia di rivelarsi limitato, o comunque di gran lunga inferiore rispetto all'omonimo regolamento Ue, sulle multinazionali della tecnologia[9].

Un altro sviluppo di rilievo nel periodo in rassegna è poi costituito dalla *International Cyberspace and Digital Policy Strategy* statunitense, pubblicata dal Dipartimento di Stato americano il **6 maggio** scorso[10]. La suddetta strategia è stata elaborata dal Dipartimento di Stato, con il supporto di altre agenzie federali, per fungere da guida nella sua azione internazionale in materia di diplomazia tecnologica e far progredire la *National Security Strategy*[11] come pure la *National Cybersecurity Strategy*[12].

La strategia è incentrata sul concetto di solidarietà digitale, ovvero la volontà di lavorare insieme su obiettivi condivisi, stare uniti, aiutare i partner a sviluppare capacità e fornire supporto reciproco. La solidarietà digitale riconosce che tutti coloro che utilizzano le tecnologie digitali nel rispetto dei diritti sono più sicuri, resilienti, autodeterminati e prosperi quando lavorano insieme per dare forma all'ambiente internazionale e reinventare l'avanguardia tecnologica. Il concetto di solidarietà digitale si basa sugli sforzi per sviluppare capacità digitali e informatiche in modo che i partner non solo siano maggiormente in grado di costruire un ecosistema digitale difendibile e resiliente nel lungo termine, ma siano anche in grado di rispondere e riprendersi rapidamente quando si verificano incidenti nonché di punirne i responsabili.

La strategia si basa su tre principi guida: 1. Una visione positiva per un cyberspazio sicuro e inclusivo fondato sul diritto internazionale, incluso il diritto internazionale dei diritti umani; 2. L'integrazione di sicurezza informatica, sviluppo sostenibile e innovazione tecnologica; 3. Un approccio politico completo che utilizzi gli strumenti appropriati della diplomazia e dell'arte di governare ("*statecraft*") sul piano internazionale nell'intero ecosistema digitale.

Al fine di creare una maggiore solidarietà digitale, la strategia individua poi le seguenti quattro aree chiave di azione: 1. Promuovere, costruire e mantenere un ecosistema digitale aperto, inclusivo, sicuro e resiliente; 2. Allineare gli approcci rispettosi dei diritti alla governance digitale e dei dati con i partner internazionali; 3. Promuovere un comportamento statale responsabile nel cyberspazio e contrastare le minacce al cyberspazio e alle infrastrutture critiche creando coalizioni e coinvolgendo i partner; 4. Rafforzare e costruire la capacità digitale e informatica dei partner internazionali.

---

[9] *Sic* M. CASTELLANETA, *Al via il primo trattato globale sull'intelligenza artificiale*, 7 giugno 2024, disponibile al link https://www.affarinternazionali.it/al-via-il-primo-trattato-globale-sullintelligenza-artificiale-2/.

[10] Cfr. USA, *International Cyberspace and Digital Policy Strategy: Towards an Innovative, Secure, and Rights-Respecting Digital Future*, 6 maggio 2024, disponibile al link https://www.state.gov/united-states-international-cyberspace-and-digital-policy-strategy/.

[11] Disponibile al link https://www.whitehouse.gov/briefing-room/statements-releases/2022/10/12/fact-sheet-the-biden-harris-administrations-national-security-strategy/.

[12] Disponibile al link https://www.whitehouse.gov/oncd/national-cybersecurity-strategy/.

Come è stato osservato, non è casuale che la suddetta strategia sia focalizzata sulla solidarietà digitale. Quest'ultima nozione, infatti, si contrappone alla sovranità digitale, caratterizzata da tendenze protezionistiche che «may give the illusion of increased control but actually pose security threats»[13]. Dette tendenze si pongono in chiara antitesi con la posizione statunitense, contraria alle disposizioni sulla localizzazione dei dati, alle tariffe per l'utilizzo della rete, alle tasse sui servizi digitali e analoghe barriere all'accesso al mercato poste in essere da Paesi autocratici quali la Russia e la Cina. Tuttavia, considerate le disparità esistenti tra gli Stati Uniti e i suoi partner, non solo nei livelli di sicurezza informatica ma anche nelle capacità, non sarà affatto facile riuscire a realizzare la solidarietà digitale a livello internazionale.

[13] Cfr. A. STEFFARO, *Building Digital Solidarity: The New International Cyberspace and Digital Policy Strategy*, 16 maggio 2024, disponibile al link https://www.centerforcybersecuritypolicy.org/insights-and-research/building-digital-solidarity-the-new-international-cyberspace-and-digital-policy-strategy.