

La (*seconda*) svolta del 2024. Anche il Consiglio d'Europa decide di regolamentare l'intelligenza artificiale

Costanza Nardocci*

THE (SECOND) 2024 TURNAROUND. EVEN THE COUNCIL OF EUROPE DECIDES TO REGULATE ARTIFICIAL INTELLIGENCE

ABSTRACT: The paper offers a critical investigation of the Council of Europe Convention on Artificial intelligence, human rights, democracy and the rule of law. It, also, aims at disclosing the novelties, similarities and differences compared to the almost contemporary Regulation of the European Union (AI Act).

KEYWORDS: Artificial Intelligence; Council of Europe; Regulation, attempt of; Human Rights; Convention.

ABSTRACT: Il saggio offre un'analisi critica della Convenzione del Consiglio d'Europa in materia di intelligenza artificiale, diritti umani, democrazia e principio di legalità. Il saggio, inoltre, si propone di mettere in evidenza le novità, differenze e somiglianze rispetto al coevo approccio dell'Unione Europea con l'AI Act.

PAROLE CHIAVE: Intelligenza artificiale; Consiglio d'Europa; Regolamentazione, tentativo di; Diritti umani; Convenzione.

SOMMARIO: 1. Considerazioni introduttive: l'accelerazione europea sulla *AI Regulation* – 2. Non solo *AI Act*: il Consiglio d'Europa e il primo trattato globale in tema di intelligenza artificiale e diritti umani – 2.1. (*Segue*) struttura e contenuti – 3. Simili ma non uguali: *ratio*, contenuti e grado di protettività, tra Trattato del CoE e *AI Act* – 4. Riflessioni conclusive sull'impatto (davvero?) globale del Trattato, oltre il Consiglio d'Europa.

«The Framework Convention on Artificial Intelligence is a first-of-its-kind, global treaty that will ensure that Artificial Intelligence upholds people's rights. It is a response to the need for an international legal standard supported by states in different continents which share the same values to harness the benefits of Artificial intelligence, while mitigating the risks. With this new treaty, we aim to ensure a responsible use of AI that respects human rights, the rule of law and democracy»¹

* *Professoressa associata in Diritto costituzionale Università di Milano. Mail: costanza.nardocci@unimi.it. Contributo sottoposto a referaggio.*

¹ Così, la Segretaria Generale del Consiglio d'Europa, Marija Pejčinović, in occasione della definitiva approvazione della Convenzione in tema di intelligenza artificiale del 17 maggio 2024.

1. Considerazioni introduttive: l'accelerazione europea sulla AI Regulation

Molto si è scritto sugli sforzi profusi nel 2023, e almeno formalmente dal 2021, dall'Unione Europea per raggiungere un accordo, poi confluito, nel marzo del 2024, nell'adozione del c.d. *AI Act*².

Meno, in parte, si è detto dei paralleli tentativi promossi dal Consiglio d'Europa che, a partire dal 2019 e tramite l'istituzione di due comitati di esperti *ad hoc* – il CAHAI³ e il CAI⁴ –, è giunto, a due mesi di distanza dalla pubblicazione dell'*AI Act*, ad approvare il primo trattato di diritto internazionale dei diritti umani in materia di intelligenza artificiale.

L'importanza del trattato del Consiglio d'Europa (CoE), la *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, va colta immediatamente dalla lista degli Stati membri e non del CoE, che hanno partecipato alle negoziazioni⁵. E l'apertura, allo stato almeno potenziale, del trattato anche oltre i confini del continente europeo⁶, suggerisce almeno tre considerazioni. La prima è che persiste una tendenza degli Stati a preferire che siano organizzazioni internazionali a disciplinare un fenomeno di difficile contenimento su scala nazionale. Ne è dimostrazione la refrattarietà con cui gli Stati membri dell'Unione Europea si sono astenuti da approvare legislazioni nazionali⁷, in attesa delle decisioni contrattate dalle istituzioni dell'Unione.

² Il testo definitivo dell'*AI Act* è consultabile al seguente link: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (ultima consultazione 02/12/2024).

³ Il CAHAI - *Ad hoc Committee on Artificial Intelligence* è stato in carica dal 2019 al 2022. In particolare, si rinvia, in questa sede, al documento con cui il CAHAI ha concluso il proprio mandato, *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, pubblicato il 3 dicembre 2021 e consultabile al seguente link: <https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d> (ultima consultazione 02/12/2024). Sull'operato del CAHAI, si veda, A. MANTELERO, *Regulating AI*, in *Beyond Data. Information Technology and Law Series*, L'Aia, 161 ss.

⁴ Il *Committee on Artificial Intelligence (CAI)* è entrato in carica nel 2022 con la finalità di redigere il successivo trattato del Consiglio d'Europa. Le attività del CAI non si sono ancora concluse ed è previsto un meeting dei membri del Comitato per il mese di settembre 2024, dove, in particolare, si discuterà della metodologia preposta alla valutazione di impatto dei sistemi di IA sui diritti fondamentali. La Draft Agenda è consultabile al seguente link: <https://rm.coe.int/cai-2024-oj2-draft-agenda/1680b0d65f> (ultima consultazione 02/12/2024).

⁵ Hanno partecipato quali Stati non membri del CoE alle negoziazioni del trattato: Stati Uniti, Canada, Messico, Giappone, Israele, Ecuador, Perù, Uruguay, Argentina e, da ultimo, anche, l'Unione Europea.

⁶ Su cui si veda § 3 dell'*Explanatory Report* sulla decisione del Comitato dei Ministri di ampliare i partecipanti alle negoziazioni del trattato. Dal Report, si legge, così, che: “[t]he Committee of Ministers also decided 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”.

⁷ In proposito, valga precisare che alcuni Stati dell'Unione Europea hanno avviato alcune iniziative a livello domestico. L'Italia ha presentato un disegno di legge in materia di IA, che, di fatto, ripete l'impostazione del Regolamento dell'Unione Europea. Si tratta del Ddl. *Schema di disegno di legge recante disposizioni e delega al Governo in materia di intelligenza artificiale*, approvato dal Governo il 23 aprile 2024. La Spagna, nel 2023, è stata istituita una apposita Agenzia indipendente, chiamata ad assolvere funzioni di supervisione sull'impiego delle tecnologie di IA, la c.d. AESIA. La Germania ha adottato un proprio *Action Plan* in tema di IA (su cui, si veda il testo al seguente link: <https://www.bmbf.de/SharedDocs/Downloads/de/2023/230823-executive-summary-ki-aktionsplan.pdf?blob=publicationFile&v=1>). Oltre i confini dell'Unione Europea può richiamarsi, in questa sede, il caso del Brasile che ha approvato, nel settembre 2023, un proprio quadro normativo in materia di IA. Il

La seconda è che simile preferenza per la dimensione sovrastatale ha spinto ordinamenti giuridici, per lungo tempo scettici nei confronti di una regolamentazione delle tecniche di intelligenza artificiale, a voltare il proprio sguardo verso chi quella direzione ha deciso, invece, di intraprendere. Il riferimento è, solo per qualche esempio, agli Stati Uniti, ma lo stesso vale per ordinamenti come quello canadese che, oltre a meritorie iniziative locali, ugualmente difetta di una disciplina federale unitaria.

Da ultimo, in terzo luogo, la centralità della *Framework Convention* va inquadrata anche in prospettiva orizzontale e non solo verticale e, cioè, nelle sue relazioni con il dato statale. Il Consiglio d'Europa è, infatti, ad oggi la prima ed unica organizzazione di diritto internazionale dei diritti umani che ha percorso la via del diritto pattizio piuttosto che appoggiarsi a strumenti di *soft law*.

Quest'ultima è stata la scelta delle Nazioni Unite, sì intervenute nel marzo del 2024 con la propria prima risoluzione in tema di intelligenza artificiale, non hanno però seguito il tracciato del Consiglio d'Europa. Inoltre, sempre allo stato attuale, non si danno iniziative analoghe da parte di sistemi di diritto internazionale regionale, primo tra tutti, di quello Inter-americano.

Il saggio si pone due obiettivi: descrivere e analizzare i contenuti del trattato; verificarne il grado di protettività, maggiore o minore, rispetto al coevo *AI Act*, nonché le eventuali possibilità ed opportunità di integrazioni reciproche rispetto all'esigenza, perseguita da entrambi i testi normativi, di assicurare un impiego delle tecnologie di intelligenza artificiale conforme ai diritti fondamentali.

2. Non solo *AI Act*: il Consiglio d'Europa e il primo trattato globale in tema di intelligenza artificiale e diritti umani

Il 17 maggio 2024, il Consiglio d'Europa ha pubblicato il testo ufficiale del proprio trattato in tema di intelligenza artificiale, la già richiamata *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*⁸.

La Convenzione costituisce l'esito di un percorso durato più di 5 anni, che ha visto la istituzione di due comitati di esperti *ad hoc* incaricati di: redigere, il primo (CAHAI), uno studio sulla fattibilità di un trattato che affrontasse compiutamente le problematiche poste dai sistemi di intelligenza artificiale in

Regno Unito, nel febbraio 2024, ha istituito l'Ufficio per l'Intelligenza Artificiale presso il Dipartimento per la Scienza, l'Innovazione e la Tecnologia ed ha adottato una propria strategia nazionale in materia dapprima nel 2021, per poi aggiornarla nel 2022. Il testo può essere letto al seguente link: https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (ultima consultazione 02/12/2024). Gli Stati Uniti, oltre alcune iniziative a livello statale particolarmente degne di nota (ci si riferisce alla recente normativa approvata dallo Stato del Colorado, CAIA, nel 2024), a livello federale ci si è arrestati al c.d. *Blueprint for an AI Bill of Rights*, consultabile al seguente link: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> (ultima consultazione 02/12/2024).

⁸ Il testo del trattato può essere letto al seguente link: <https://rm.coe.int/1680afae3c>. A commento del testo, si vedano i contributi di A. HARS, *Conceptual Difficulties in the Transformation of Human Rights to the Realm of Artificial Intelligence*, in *Acta Humana*, 2, 2024, 123 ss.; F.P. LEVANTINO, F. PAOLUCCI, *Advancing The Protection Of Fundamental Rights Through AI Regulation: How The Eu And The Council Of Europe Are Shaping The Future*, in P. CZECH, L. HESCHL, K. LUKAS, M. NOWAK, G. OBERLEITNER (a cura di), *European Yearbook on Human Rights 2024*, 2025, Leiden, in corso di pubblicazione e consultabile al seguente link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4881656 (ultima consultazione 02/12/2024).

relazione ai valori del Consiglio d'Europa⁹; il secondo (CAI) di avviare e condurre le negoziazioni tra Stati membri e *stakeholders* esterni funzionali alla redazione dei contenuti del futuro trattato¹⁰.

Rinviando al paragrafo successivo una riflessione sui punti di contatto, ma anche, soprattutto, di divergenza tra il testo in esame e l'*AI Act*, pare opportuno mettere qualche punto fermo sui contenuti della Convenzione e sulle finalità che il Consiglio d'Europa si è proposto di soddisfare tramite la sua adozione.

Un primo aspetto da sottolineare di più ampio respiro concerne la chiara scelta di campo del Consiglio d'Europa che precede le Nazioni Unite e altre organizzazioni regionali, nel segno della regolamentazione, anche se di principio, dell'impiego dei sistemi di intelligenza artificiale nello spazio europeo.

Ancora, altrettanto chiara è la *ratio* che sottende al trattato e, cioè, la definizione precisa del quadro di principi o criteri guida, che dovranno orientare i programmatori e gli utilizzatori dei sistemi di intelligenza artificiale in tutte le fasi del loro sviluppo¹¹.

Se non sono sottovalutate le potenzialità dei sistemi di intelligenza artificiale, è però altrettanto chiaro già nel Preambolo, riletto alla luce del Rapporto Esplicativo, che il Consiglio d'Europa riconosce alcune priorità sul piano dell'impatto delle tecnologie di intelligenza artificiale su alcuni principi fondamentali, identificati con estrema puntualità sin dalle prime righe del testo convenzionale.

2.1. (Segue) struttura e contenuti

Venendo ai contenuti del testo, è l'ampiezza di scopo che costituisce un primo tratto caratterizzante della Convenzione.

L'Articolo 1 si inserisce in questa prospettiva, precisando che le norme del testo convenzionale mirano ad assicurare il rispetto dei diritti umani, del principio democratico e di legalità da parte di tutti i sistemi di intelligenza artificiale in ogni fase del loro c.d. "*lifecycle*".

Questa apertura all'applicazione del testo della Convenzione a tutte le tecnologie di intelligenza artificiale, senza distinzioni, viene poi meglio precisata al paragrafo seguente che risponde alle esigenze, mutevoli e assai dinamiche, dello sviluppo della tecnologia in esame.

Ne discende che, nelle intenzioni degli estensori, le norme convenzionali andranno ad agire secondo un'intensità variabile in dipendenza dalle specificità del sistema e del suo impatto, più o meno severo, sui diritti fondamentali salvaguardati, anzitutto ma non solo, a livello del Consiglio d'Europa.

⁹ In particolare, si veda CAHAI, *Possible elements of a legal framework on artificial intelligence, based on Council of Europe's standards on human rights, democracy and the rule of law*, adottato nel Dicembre 2021.

¹⁰ Alle negoziazioni hanno, infatti, partecipato anche 68 NGOs insieme ad altre organizzazioni internazionali tra cui: l'Organisation for Security and Co-operation in Europe (OSCE), l'Organisation for Economic Co-operation and Development (OECD), la United Nations Educational, Scientific and Cultural Organisation (UNESCO) e altri organismi operanti in seno al Consiglio d'Europa. Così, precisa l'*Explanatory Report*, § 4.

¹¹ Chiaro l'*Explanatory Report*, che precisa che: "[The] Framework Convention focuses on the protection and furtherance of human rights, democracy and the rule of law, and does not expressly regulate the economic and market aspects of artificial intelligence systems. Taken as a whole, it provides a common legal framework at the global level in order to apply the existing international and domestic legal obligations that are applicable to each Party in the sphere of human rights, democracy and the rule of law of each Party and aims to ensure that the activities within the lifecycle of artificial intelligence systems by both public and private actors comply with these obligations, standards and commitments".

Espressione di simile impostazione, di ampio respiro, è anche la definizione di intelligenza artificiale recepita dal testo del trattato e codificata ai sensi dell'art. 2. Si legge, così, che, per intelligenza artificiale, la Convenzione Quadro copre ogni: «*machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments*»¹².

È il Rapporto Esplicativo che chiarisce le ragioni della scelta: collega la nozione di intelligenza artificiale recepita dal testo convenzionale a quella dell'OECD¹³; e ne circoscrive, poi, i confini, sì da escluderne la riferibilità a tecnologie il cui funzionamento dipenda solo da regole stabilite dal programmatore e che si limitino all'assolvimento di mansioni esecutive. Una nozione, però, che il testo rende volutamente astratta e flessibile, allo scopo di assicurarne la c.d. neutralità tecnologica e, quindi, la ultrattività del trattato¹⁴.

Chiude, infine, il Capo I la norma dedicata all'ambito applicativo del trattato, l'art. 3.

L'aspetto più rilevante va colto in quanto stabilito ai sensi della lettera *b*) del paragrafo 1, dove la Convenzione richiede agli Stati contraenti la stesura di una dichiarazione ("*declaration*") che dettagli le modalità ("*how*") tramite le quali lo Stato intende implementare le obbligazioni del trattato¹⁵.

Accanto alle obbligazioni generali di cui al *Chapter II*, tra gli aspetti più significativi del trattato vi è, però, senza dubbio la scelta di delineare con cura i principi, i relativi corollari e i criteri guida a cui sono chiamati a sottostare gli Stati contraenti in tutte le fasi della vita dei sistemi di intelligenza artificiale e, cioè, la progettazione, l'implementazione, e, da ultimo, l'impiego dei sistemi di intelligenza artificiale da parte dell'utilizzatore finale. Il *Chapter III* del testo convenzionale affianca, cioè, a diritti tradizionali del diritto pubblico, costituzionale e sovranazionale, principi "nuovi" oppure, se così si preferisse definirli, criteri guida che sono venuti affermandosi parallelamente alla discesa delle tecnologie di intelligenza artificiale nello spazio "umano", pubblico e privato.

Gli articoli da 6 a 13 compresi dettagliano, quindi, un insieme di principi che, nelle intenzioni degli estensori del trattato, dovranno governare le multiformi relazioni tra persona umana e intelligenza artificiale. Tra questi, meritano di essere sinteticamente richiamati: la dignità umana, a cui la Convenzione affianca, non così sorprendentemente, il diritto alla c.d. "autonomia individuale"¹⁶; il principio di

¹² Così l'art. 2.

¹³ Sulla definizione di IA, è interessante richiamare l'*Explanatory Memorandum on the updated OECD definition of an AI system*, pubblicato nel marzo del 2024, consultabile al seguente link: <https://www.oecd-ilibrary.org/doc-server/623da898-en.pdf?expires=1722520019&id=id&accname=guest&checksum=71863626A100B8ADF8E04D28B7EBEC7C> (ultima consultazione 02/12/2024). In particolare, si richiama un passaggio del testo che investe l'applicazione della definizione proposta di IA, di cui viene enfatizzata la voluta ampiezza. Si legge, così, che: «[t]he updated definition of AI is inclusive and encompasses systems ranging from simple to complex. AI represents a set of technologies and techniques applicable to many different situations. Specific techniques, such as machine learning, may raise particular considerations for policymakers, such as bias, transparency, and explainability, and some contexts of use (e.g., decisions about public benefits) may raise more significant concerns than others. For that reason, when applied in practice, additional criteria may be needed to narrow or otherwise tailor the definition when used in a specific context».

¹⁴ Si veda, per questo, aspetto, il par. 24 dell'*Explanatory Report*.

¹⁵ Sempre l'art. 3 esclude, che le obbligazioni scaturenti dal trattato possano in qualsiasi modo interferire in materia di sicurezza nazionale.

¹⁶ Cfr. art. 7.

trasparenza e di controllo umano sul funzionamento delle tecnologie di intelligenza artificiale (il c.d. *oversight*)¹⁷; il principio di responsabilità (*accountability*)¹⁸; i principi, classici, di eguaglianza e di non discriminazione¹⁹, di riservatezza e di protezione dei dati²⁰, per poi chiudere con la c.d. *reliability*²¹, cioè la robustezza e affidabilità dei sistemi e la sicurezza (*safe innovation*)²².

Ciascuno di questi principi meriterebbe un approfondimento a sé stante. In questa sede, ci si limiterà, tuttavia, ad alcune osservazioni sui tratti più caratterizzanti e specifici di ciascuno. L'intento è, soprattutto, evidenziare gli aspetti che più direttamente investono le relazioni tra la "macchina" e l'"umano" e che la Convenzione ha tradotto enucleando la lista di principi sopra tratteggiata.

In questo quadro, primario rilievo riveste la nozione di "autonomia individuale", che il Rapporto Esplicativo definisce corollario del principio di autodeterminazione, enfatizzandone, però, al tempo stesso, i significati più specifici assegnategli nel contesto specifico dell'intelligenza artificiale. Autonomia individuale è, allora, controllo umano sull'operato della macchina, che non deve mai operare finendo con il comprimere le volontà e l'*agere* della persona.

E il Rapporto Esplicativo è ancora più esplicito. La salvaguardia dell'autonomia individuale, si legge, è infatti tanto più importante al cospetto di tecnologie dotate di elevate abilità di imitazione e di manipolazione. Una rilettura del diritto all'autodeterminazione individuale ai "tempi" dell'intelligenza artificiale, che si rivela in definitiva preziosa anche nella prospettiva dell'operato di giudici e Corti, il cui intervento si prospetta sempre più frequente nei prossimi anni.

Il Capo prosegue, poi, occupandosi dei principi di trasparenza e di controllo, cioè di *oversight* sul funzionamento dei sistemi di intelligenza artificiale in tutte le fasi che presiedono il loro impiego.

L'interpretazione autentica di questi due principi, che si ricava ancora una volta dal Rapporto Esplicativo, concorre a posizionare la c.d. *transparency* e lo *human oversight* al centro del reticolo di principi salvaguardati dal trattato. In una nota a piè di pagina del testo, gli estensori della Convenzione hanno, così, la cura di appuntare l'attenzione sullo stretto legame che deve riconoscersi alle relazioni tra i

¹⁷ Cfr. art. 8.

¹⁸ Cfr. art. 9.

¹⁹ Cfr. art. 10. In letteratura, sulle relazioni tra IA e discriminazioni, si vedano F. ZUIDERVEEN BURGESIUS, *Discrimination, artificial intelligence, and algorithmic decision-making*, Strasburgo, 2018; J. KLEINBERG, J. LUDWIG, S. MULLAINATHAN, C.R. SUNSTEIN, *Discrimination in the Age of Algorithms*, in *Journal of Legal Analysis*, 10, 2018, 113 ss.; J. KLEINBERG, J. LUDWIG, S. MULLAINATHAN, A. RAMBACHAN, *Algorithmic fairness*, in *AEA papers and proceedings*, 108, 2018, 22 ss.; J. GERARDS, R. XENIDIS, *Algorithmic discrimination in Europe: Challenges and Opportunities for EU equality law*, 3 dicembre 2020, in <https://www.europeanfutures.ed.ac.uk/algorithmic-discrimination-in-europe-challenges-and-opportunities-for-eu-equality-law/> (ultima consultazione 02/12/2024); P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, in AA.Vv. (a cura di), *Liber amicorum per Pasquale Costanzo – Diritto Costituzionale in trasformazione Vol. I – Costituzionalismo, Reti e Intelligenza artificiale*, Genova, 2020, 457 ss.; si consenta, infine, il rinvio a C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, in *Rivista "Gruppo di Pisa"*, 3, 2021, 9-60; in relazione alle specificità delle interferenze tra IA, genere e discriminazioni, si vedano M. D'AMICO, C. NARDOCCI, *Discriminazioni, Donne e Intelligenza Artificiale*, in G. CERRINA FERONI, C. FONTANA, E.C. RAFFIOTTA (a cura di), *AI Anthology*, Bologna, Il Mulino, 2022.

²⁰ Cfr. Articolo 11.

²¹ Cfr. Articolo 12.

²² Cfr. Articolo 13. In dottrina, per un approfondimento sul tema dell'affidabilità dei sistemi di IA in sede di applicazione giudiziaria, si veda S. PENASA, *Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale in prospettiva comparata*, in *DPCE online*, 1, 2022, 297 ss.

principi di trasparenza e *oversight*, da un lato, e l'autonomia e autodeterminazione individuale, reinterpretate alla luce delle implicazioni derivanti dai nuovi sistemi di intelligenza artificiale, dall'altro. I contenuti della norma convenzionale sono, infatti, ulteriormente precisati dal Rapporto Esplicativo, che considera partitamente due aspetti: la spiegabilità del sistema di intelligenza artificiale (*explainability*), intesa come capacità del sistema di intelligenza artificiale in esame di fornire spiegazioni sufficienti circa le proprie modalità di funzionamento, nonché delle finalità del suo utilizzo²³; e l'interpretabilità (*interpretability*), che richiama, viceversa, la possibilità, dall'esterno, di comprendere come la tecnologia di intelligenza artificiale compie le proprie decisioni e previsioni²⁴.

Agli obblighi di trasparenza e di *oversight* fanno, poi, seguito i principi di cui all'art. 9 e, cioè, la c.d. *accountability* e la responsabilità. *Accountability* e responsabilità, che dovrebbero assicurare la identificazione del soggetto, individuale oppure collettivo, chiamato a rispondere in ipotesi di effetti del sistema di intelligenza artificiale lesivi dei diritti umani.

Il Rapporto Esplicativo, come sempre, si sofferma sui contenuti della norma, mettendo in rilievo due aspetti²⁵.

Il primo inserisce i principi in esame all'interno del più ampio discorso sulla *liability* applicata alle tecnologie di intelligenza artificiale, evidenziando l'esigenza (o, forse, l'urgenza), che sussista sempre la possibilità di offrire risposte adeguate, in termini di individuazione del soggetto responsabile, in relazione a ciascuna delle fasi in cui si snoda la vita del sistema di IA considerato. Ciò comporta il recepimento di una concezione ampia e flessibile di responsabilità, che non guarda allo status del programmatore, di colui che implementa il sistema e, infine, dell'utilizzatore, diversificandone a monte e in modo astratto le responsabilità. Piuttosto, si impone agli Stati di garantire, tramite l'adozione di meccanismi *ad hoc*, che sia sempre accertata e accertabile la responsabilità del privato oppure del pubblico in costanza di una violazione dei diritti umani, della democrazia, del principio di legalità.

Il secondo punto attiene, invece, alle relazioni tra la responsabilità, da una parte, e la trasparenza e il controllo umano (*oversight*), dall'altra. Il Rapporto Esplicativo non manca, cioè, di sottolineare come i secondi siano strumentali alla prima: tanto più saranno garantite la spiegabilità e la interpretabilità del sistema di intelligenza artificiale, tanto più agevole si rivelerà, a sua volta, la comprensione della fase della vita della "macchina" a cui addebitare l'effetto lesivo dei diritti fondamentali e, con essa, il soggetto, umano, chiamato a risponderne²⁶.

Quest'ultimo passaggio è particolarmente significativo nell'impianto della Convenzione.

Se letto unitariamente alla norma che impone la valutazione del rischio del sistema di IA in termini di impatto sui diritti umani, i principi in esame appaiono tanto più importanti nel supportare lo sviluppo di strategie di *assessment* che guardino all'intero sviluppo della tecnologia di IA al fine di distinguere fasi e sotto-fasi strumentali alla contestuale identificazione del soggetto o dei soggetti responsabili in relazione a ciascuna. In questo senso, muove in modo esplicito anche la *ratio* della Convenzione, che

²³ Cfr. § 60 del Rapporto Esplicativo.

²⁴ *Ivi*, Cfr. § 61.

²⁵ Cfr. § 66 ss.

²⁶ Cfr. § 69.

pone in relazione l'art. 9 con le obbligazioni positive che scaturiscono dall'art. 16 in materia di valutazione del rischio delle tecnologie di IA²⁷.

Si inserisce, invece, nel novero dei principi tradizionali del diritto pubblico la salvaguardia dell'eguaglianza e la non discriminazione di cui riferisce l'art. 10 del trattato. Qui, il profilo meritorio attiene alla puntuale delineazione delle cause che, nel quadro delle interazioni tra macchina e persona, sono suscettibili di tradursi in trattamenti irragionevolmente diversificati tra individui e gruppi. Il Rapporto Esplicativo elenca diverse tipologie di *bias*²⁸, appuntando l'attenzione sulle scansioni temporali della vita del sistema di intelligenza artificiale che possono originare lesioni dei principi di eguaglianza e di non discriminazione. Aspetto di cui si dirà più diffusamente nel paragrafo che segue e che separa l'impostazione euro-unitaria dall'approccio degli estensori del testo convenzionale.

In senso analogo, è anche la disposizione di cui al paragrafo 2, che rimette agli Stati la valutazione circa l'adeguatezza o meno delle normative domestiche vigenti per contrastare la discriminazione originata dall'impiego delle tecnologie di IA.

La Convenzione prosegue, poi, occupandosi di un altro diritto classico, la riservatezza e la protezione dei dati. Enfasi è riposta sull'ovvia centralità che la *privacy* e la tutela della segretezza rivestono nello specifico contesto esaminato e il Rapporto Esplicativo – questo il tratto più interessante – incoraggia gli Stati contraenti ad assicurare quel *surplus* di garanzie attivabili a livello nazionale così come a garantire il coordinamento con la normativa vigente in materia di diritto dell'Unione Europa con un richiamo esplicito al Regolamento sulla protezione dei dati, il c.d. GDPR²⁹.

Questa parentesi, che aggancia la Convenzione a due dei diritti classici del diritto pubblico maggiormente esposti al rischio di essere oggetto di violazione per effetto del funzionamento dei sistemi di intelligenza artificiale, è chiusa dal riferimento ad un principio, viceversa, "nuovo" o, si potrebbe dire, "AI-specific".

²⁷ Cfr. § 70. In tema di valutazione del rischio dei sistemi di AI, la prima proposta degna di nota a livello sovranazionale risale al Joint Technical Committee ('JTC') dell'International Organization for Standardization (ISO) e dell'International Electro technical Commission (IEC), il cui lavoro è sfociato nella pubblicazione dei c.d. standard ISO, *Information Technology—Artificial Intelligence—Artificial Concepts and Terminology*, ISO/IEC 22989:2022(E), luglio 2022 ('ISO/IEC 22989'). A commento, si rinvia al contributo di J.M. BELLO Y VILLARINO, *Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI?*, in E. SHIRLOW, D.R. ROTHWELL (a cura di), *The Australian Year Book of International Law Online*, Leiden, 2023, 157 ss.

²⁸ Cfr. § 75. In particolare, il Rapporto richiama i seguenti: "potential bias of the algorithm's developers; [...] potential bias built into the model upon which the systems are built; [...] potential biases inherent in the training data sets used [...], or in the aggregation or evaluation of data; [...] biases introduced when such systems are implemented in real world settings [...]; automation or confirmation biases, whereby humans may place unjustified trust in machines and technological artefacts or situations where they select information that supports their own views, in both cases ignoring their own potentially contradictory judgment and validating algorithmic outputs without questioning them [...]". Si tratta, da prospettiva differente, dell'insieme delle c.d. fasi in cui può annidarsi il rischio di una deriva discriminatoria del sistema di IA. In letteratura, su questo aspetto, si vedano, diffusamente, S. BAROCAS, A.D. SELBST, *Big data disparate impact*, in *California Law Review*, 104, 2016, 671 ss.

²⁹ In tema, per un approfondimento sulle relazioni tra intelligenza artificiale e GDPR, si rinvia a G. SARTOR, F. LAGIOIA, *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence (study prepared for the European Parliament)*, Bruxelles, 2020, 15 ss., in [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf) (ultima consultazione 02/12/2024).

L'art. 12 tratta, cioè, della c.d. *reliability* o affidabilità delle tecnologie di intelligenza artificiale. La norma impone alle Parti contraenti l'obbligo di mettere a punto meccanismi capaci di salvaguardare i corollari del principio in esame: la robustezza, la sicurezza, l'integrità e la protezione dei dati, la sicurezza *cyber*. Ancora una volta, il tema è l'adozione di *standard* che governino il funzionamento dei sistemi di IA sì da certificarne un *agere* conforme al principio di affidabilità. Il Rapporto Esplicativo, però, va oltre la delineazione di una obbligazione generica di risultato e dice qualcosa di più. Arriva, così, a suggerire che gli Stati contraenti si dotino di enti, cioè attribuiscono a *stakeholders* esterni, il compito di verificare la efficacia delle strategie messe a punto dal programmatore, dallo sviluppatore, dall'utilizzatore finale, singolarmente oppure unitariamente considerati, allo scopo di evitare una garanzia solo fallace delle esigenze di affidabilità che, peraltro, evidentemente presentano una forte connessione con i principi di trasparenza e di *oversight*, nonché con la responsabilità di cui, rispettivamente, agli artt. 8 e 9 del trattato³⁰.

Chiude, infine, il Capo III, la norma dedicata alla "*safe innovation*". Si tratta – lo ribadisce il Rapporto Esplicativo³¹ – del principio che, più di tutti, tocca al cuore le dinamiche esistenti tra impiego dell'intelligenza artificiale nella dimensione pubblica e privata, i diritti umani e la democrazia. Il Rapporto Esplicativo non si astiene dall'offrire indicazioni e proposte funzionali a promuovere uno sviluppo dell'innovazione tecnologica che non si traduca in violazione dei principi fondanti lo Stato di diritto. Tra le molte, vi è il richiamo alle "*regulatory sandboxes*", ma, anche, a linee guida specifiche che possano orientare tutti i soggetti coinvolti nella disciplina delle molte fasi di vita della macchina.

La Convenzione tratta, poi, dei rimedi di fronte a violazioni dei diritti fondamentali.

Se l'art. 14 chiama gli Stati ad applicare e a rileggere i meccanismi già vigenti a livello nazionale e di diritto internazionale dei diritti umani, l'art. 15 tratteggia una obbligazione specifica alla luce delle specificità di contesto sottostante all'impiego dei sistemi di intelligenza artificiale e alle loro ricadute sui diritti della persona. La norma stabilisce l'opportunità di disporre garanzie procedurali di *oversight*, operanti sia *ex ante* che *ex post*, in base alle quali vagliare il corretto e non pregiudizievole utilizzo delle tecnologie di IA. Il paragrafo secondo è particolarmente dettagliato e, così, il Rapporto Esplicativo che indugia sugli esempi di meccanismi da implementare a livello domestico in conformità del trattato³². Sicuramente centrale nell'impianto della Convenzione del Consiglio d'Europa non è solo la preferenza per un apparato rimediabile, che integri gli strumenti già esistenti e "generalisti" esistenti a livello nazionale e sovranazionale con altri, ancora una volta, "*AI-specific*". Coerente con l'approccio dell'*AI Act* è, così, il Capo V dedicato alla verifica della ricorrenza e, in caso positivo, al contenimento dei rischi derivanti dal ricorso ai sistemi di intelligenza artificiale.

Alla luce dei principi delineati dal Capo III, l'art. 16, rubricato "*Risk and impact management framework*", prevede che gli Stati contraenti si dotino di meccanismi in grado di identificare, valutare,

³⁰ Cfr. § 84 ss.

³¹ Cfr. § 90 ss.

³² Cfr. § 104. Interessa richiamare l'attenzione con cui il Rapporto Esplicativo si preoccupa di chiarire gli obblighi degli Stati contraenti ogniqualvolta sussista un'interazione tra la persona umana e la "macchina", precisando che: "persons interacting with an artificial intelligence system should be duly notified that they are interacting with an artificial intelligence system rather than with a human". Si tratta di una affermazione che si ricollega evidentemente al principio del c.d. *right to know*, qui declinato come diritto di chi sia soggetto al funzionamento di dette tecnologie e non di chi faccia uso delle medesime.

prevenire e contenere i rischi scaturenti dall'impiego delle tecnologie di intelligenza artificiale valutandone l'impatto, anche potenziale, sui diritti umani, sul principio democratico e di legalità. Il paragrafo 2 suggerisce alcune strategie di intervento tra cui merita richiamare la possibilità di testare *ex ante* il sistema di IA prima della sua messa in funzione.

La norma non circoscrive la valutazione di impatto a certe tipologie di sistemi di intelligenza artificiale e rimette alla discrezionalità delle Parti contraenti l'eventuale introduzione di moratorie oppure di divieti laddove talune tecnologie si rilevino particolarmente lesivi dei diritti fondamentali³³.

Chiudono la Convenzione i Capi VI, VII e VIII, dedicati, rispettivamente, alla implementazione del trattato, ai meccanismi di cooperazione tra gli Stati contraenti sino alle clausole finali, a cui si dedicherà qualche riflessione in chiusura del presente paragrafo soprattutto con riferimento alle norme in tema di ratifica, riserve e risoluzione delle controversie relative all'applicazione del trattato.

Per quanto attiene al Capo VI, la Convenzione Quadro ha la cura di precisare che l'implementazione del trattato: non può, in alcun caso, risolversi in una violazione dei principi di eguaglianza e di non discriminazione; deve tenere conto delle esigenze dei minori e delle persone con disabilità, favorire l'acquisizione di competenze informatiche da parte di tutti e tutte, non pregiudicare *surplus* di tutela garantiti dalle legislazioni nazionali, né comprimere diritti salvaguardati a livello domestico³⁴.

Interessante è la previsione di cui all'art. 19, "*Public consultation*", che impone agli Stati parte di assicurare il coinvolgimento, tramite consultazioni pubbliche, di *stakeholders* esterni in ogni aspetto che attenga alle implicazioni sociali, economiche, giuridiche, etiche, ambientali dei sistemi di intelligenza artificiale.

Si tratta di norma da guardare con particolare favore per almeno due ragioni.

La prima è che la scelta del Consiglio d'Europa di incoraggiare scambi con associazioni non governative e altri enti, pubblici e privati, recepisce una prassi che è emersa nel corso delle negoziazioni dell'*AI Act*. Al di là di come il Regolamento dell'Unione Europea abbia, poi, accolto o meno i suggerimenti ricevuti, è però evidente che, dal 2021 al 2024, il testo è stato oggetto di alcuni importanti *Statement* che ne hanno sottolineato fragilità e punti critici³⁵. La seconda ragione, connessa alla prima, si coglie nel ruolo delle associazioni non governative, che il trattato valorizza, mettendo a sistema l'esigenza di supportare la partecipazione "dal basso" nel processo che guarda alla implementazione delle disposizioni convenzionali.

Sui meccanismi di cooperazione, ci si limita, in questa sede, a riprendere quanto stabilisce l'art. 23, che invita gli Stati contraenti a riunirsi periodicamente in un'apposita Conferenza allo scopo, tra gli altri, di discutere della implementazione della Convenzione, della opportunità di apporvi emendamenti, di definire eventuali controversie tra le Parti, nonché di favorire lo scambio con *stakeholders* esterni.

Oltre alla necessaria cooperazione tra gli Stati, la Convenzione, seguendo in questo l'impostazione euro-unitaria, non manca di sollecitare l'istituzione di appositi meccanismi di controllo e di *compliance* delle norme del trattato, secondo quanto stabilisce l'art. 26. Meccanismi di controllo, che –

³³ Sul carattere eterogeneo della norma rispetto all'art. 27 dell'*AI Act*, si veda, *infra*, il paragrafo successivo.

³⁴ Così l'art. 18.

³⁵ A commento della prima versione del testo, si rinvia, diffusamente, a M. VEALE, F. ZUIDERVEEN BORGESIU, *Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, in *Computer Law Review International*, 4, 2021, 105 ss.

precisa il paragrafo 2 – devono rispondere ai canoni di indipendenza e di imparzialità, così come lo Stato dovrà dotare tali enti di poteri adeguati a rispondere in modo efficiente alle funzioni assegnategli. La Convenzione Quadro non impone la creazione di enti *ad hoc* quale unica strategia per rispondere agli obblighi derivanti dal trattato. Tuttavia, laddove decida di avvalersi di enti di previa istituzione oppure di articolare l'apparato di supervisione in una molteplicità di organismi, lo Stato parte è chiamato a disciplinare che tali enti non operino in modo frammentato, assicurandone il coordinamento. Da ultimo, si collocano le disposizioni finali, che contengono qualche punto meritevole di sintetica notazione.

Anzitutto, la Convenzione è stata aperta alla ratifica anche da parte di Stati non membri del Consiglio d'Europa. Lo prevede esplicitamente l'art. 31, che dispone che il Comitato dei Ministri potrà invitare Stati non membri del Consiglio d'Europa, che non abbiano preso parte alle negoziazioni, a firmare e ratificare il trattato secondo quanto stabilito a norma dell'art. 20, lett. d) dello Statuto del Consiglio d'Europa³⁶.

In secondo luogo, e venendo alla sua forza precettiva, il trattato non potrà essere fatto oggetto di riserva da parte degli Stati contraenti, escludendo, così, compressioni o limitazioni dell'ambito applicativo della Convenzione. Una scelta meritoria soprattutto, perché, così facendo, il Consiglio d'Europa blinda il Capo III, sancendo l'inderogabilità per tutti gli Stati parte dei principi *ivi* contenuti.

Seguendo la tradizionale impostazione dei trattati di diritto internazionale e per concludere, anche la *Framework Convention* disciplina l'istituto della denuncia del trattato e, quindi, la possibilità per lo Stato parte di sottrarsi alle obbligazioni a cui aveva acconsentito di vincolarsi quale effetto della ratifica.

3. Simili ma non uguali: ratio, contenuti e grado di protettività tra Trattato del CoE e AI Act

Oltre l'analisi delle previsioni del trattato, interessa interrogarsi sulle relazioni intercorrenti tra la *Framework Convention* del Consiglio d'Europa e il Regolamento dell'Unione Europea, il già più volte evocato *AI Act*, secondo una prospettiva che ne evidenzia frizioni, se ve ne sono, oppure contenuti non del tutto sovrapponibili se non, addirittura, divergenze.

Si tratta di un tema di sicuro rilievo se si considerano due dati fattuali preliminari. Il primo è che, dal punto di vista temporale, le attività del Consiglio d'Europa si sono svolte quasi parallelamente al dibattito sviluppatosi in seno alle istituzioni dell'Unione Europea, cosa che avrebbe potuto propendere per una influenza biunivoca tra le due organizzazioni internazionali. Il secondo concerne, invece, la partecipazione dell'Unione Europea alle negoziazioni della *Framework Convention*. Elementi, che avrebbero potuto suggerire una maggiore comunanza di intenti sin dall'inizio tra le organizzazioni internazionali e che, ad oggi, si è tradotta nell'adozione di due testi che, forse, costituisce sì l'esito di un'influenza reciproca che, però, ha portato ad esiti più positivi per l'*AI Act* di quanto non si sia invece riscontrato in relazione al trattato del Consiglio d'Europa³⁷.

³⁶ Il riferimento è alla maggioranza qualificata dei 2/3, a cui dovrà, eventualmente, sottostare l'approvazione della possibilità che Stati non membri del Consiglio d'Europa possano firmare e ratificare la Convenzione Quadro.

³⁷ Non a caso, il testo definitivo del trattato è stato fatto oggetto di critiche da parte di un gruppo di associazioni non governative, che hanno reso pubbliche le proprie perplessità in una *open letter* pubblicata il 5 marzo 2024.

Sebbene le divergenze di *ratio* e di impostazione fossero inizialmente più marcate – l’Unione Europea più incline ad occuparsi di una regolamentazione sufficientemente protettiva del diritto alla riservatezza e alla protezione dei dati ed un Consiglio d’Europa più attento alle ricadute delle tecnologie di IA sui diritti individuali –, persistono alcuni elementi che separano il Consiglio d’Europa dall’Unione Europea nel rispettivo approccio alla regolamentazione dell’intelligenza artificiale.

Alcune scelte, in particolare quelle che insistono sugli aspetti che hanno occupato una posizione di primo piano nel dibattito precedente all’approvazione dell’AI Act nel marzo del 2024, hanno per lungo tempo caratterizzato l’operato del Consiglio d’Europa, sino alla pubblicazione del *Consolidated Working Draft* nel luglio del 2023.

Inizialmente divergenti rispetto alla direzione intrapresa dall’Unione Europea, alcune soluzioni proposte dal Consiglio d’Europa si sono in seguito ricomposte, in occasione della stesura definitiva del testo del trattato che si è progressivamente avvicinato all’AI Act. Altre, che, originariamente divergenze non erano ma che, anzi, esprimevano un’impostazione condivisa tra le due organizzazioni internazionali, sono state invece, per così dire, “assorbite”, cioè rese non più esplicite dal trattato, tanto che, ad oggi, non ne è del tutto pacifica l’interpretazione in termini di omogeneità oppure di eterogeneità rispetto alle opzioni accolte dal legislatore dell’Unione.

Un esempio delle prime, cioè delle c.d. “divergenze originarie” poi risolte, è costituita dalla definizione di intelligenza artificiale. Se quella contenuta nelle versione vigente del trattato è perfettamente coincidente con la nozione, più ristretta, di cui all’AI Act³⁸, non così è stato almeno sino al 2024. La

Il testo integrale può essere consultato al seguente link: https://docs.google.com/document/d/19pwQgOr7g5Dm6_OIRvTAgBPGXaufZrNW/edit (ultima consultazione 02/12/2024).

³⁸ Anche in seno all’AI Act, la definizione di intelligenza artificiale è stata protagonista di vicende alterne. Si è passati da una nozione ampia di intelligenza artificiale, secondo cui un sistema di IA equivarrebbe a qualsiasi: «software that is developed with one or more of the techniques and approaches listed in [the AI Act] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with»; definizione criticata per la sua incapacità di tracciare una distinzione netta tra semplice *softwares* e sistemi di IA veri e propri. A questa prima opzione, ne è seguita una seconda, più ristretta, suggerita dal Consiglio dell’Unione Europea, che così qualificava le tecnologie di IA: «[AI system] means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts». L’approccio del Consiglio, poi più in linea con la definizione conclusiva, mirava a circoscrivere la nozione di IA ai sistemi di *machine learning, logic, o knowledge-based*. La definizione più circoscritta è stata oggetto di critiche, in modo particolare, da alcune associazioni non governative, tra cui si richiama, in questa sede, *Algorithmwatch* e *AccessNow*, nel documento *EU policy makers: Protect people’s rights, don’t narrow down the scope of the AI Act!*, 23 novembre 2021. Valga ricordare che, almeno originariamente, la definizione di partenza e di riferimento per le azioni successive tanto dell’Unione Europea quanto del Consiglio d’Europa era stata proposta dall’*Independent High-Level Expert Group on Artificial Intelligence* nel 2019, che così si esprimeva: «[w]e propose to use the following updated definition of AI: Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions».

definizione del *Consolidated Working Draft* era, infatti, più ampia e rispondeva in modo più convincente al tema dei tempi dell'innovazione tecnologica, che non sono i tempi del diritto, ammettendo interpretazioni oppure aggiustamenti della nozione di intelligenza artificiale proposti dagli Stati parte proprio con la finalità di rispondere tempestivamente al dinamismo delle nuove tecnologie. Accanto a questo aspetto, se da più parti è stata per lungo tempo messa in discussione la decisione dell'Unione Europea di abbracciare una nozione ristretta di intelligenza artificiale, circoscritta pressoché soltanto ai sistemi di *machine learning*, e che si scontrava con l'impostazione viceversa più onnicomprensiva del *Consolidated Working Draft*, oggi tale frizione non esiste più con il Consiglio d'Europa, come già detto, che ha fatto un passo indietro negli ultimi mesi ripiegando sulla stessa proposta definitoria dell'Unione.

Tre le seconde – casi, cioè, di convergenza tra le due organizzazioni sfumati nelle versioni definitive dei testi –, si inserisce l'opportunità o meno di classificare le tecnologie di IA in base al criterio del rischio, chiarendo che cosa si intenda per rischio.

Se l'*AI Act*, condivisibilmente, chiarisce che cosa costituisce un "rischio" nella prospettiva della regolamentazione dell'Unione Europea³⁹, il Consiglio d'Europa ha intrapreso una strada parzialmente differente. Nonostante la *ratio* del trattato sia sovrapponibile a quella dell'*AI Act*, nel senso di guardare alla legittimità del ricorso ai sistemi di intelligenza artificiale soltanto laddove non costituiscano un rischio per la tenuta dei diritti individuali, la *Framework Convention* non menziona più, come viceversa faceva il testo del *Consolidated Working Draft*, la propria adesione al criterio del rischio. Il *Consolidated Working Draft*, ancora una volta secondo una impostazione preferibile perché meglio delineava i requisiti a cui subordinare l'impiego dei sistemi di IA, dedicava infatti una disposizione apposita al c.d. criterio del rischio⁴⁰. Disposizione che, viceversa, scompare nel testo del trattato, che, peraltro, non segue nemmeno l'impostazione finale dell'*AI Act*, omettendo qualsiasi riferimento esplicito anche soltanto alla definizione della nozione di "rischio".

I due esempi sono interessanti, perché evidenziano come la contaminazione tra le due organizzazioni che, per prime, hanno abbracciato un approccio favorevole alla *regulation* dei sistemi di IA si sia poi tradotta in esiti non del tutto conformi, sebbene uno scambio tra Unione Europea e Consiglio d'Europa sia più che evidente, nel bene e nel male⁴¹. Nel male, se si guarda alla nozione più ristretta di intelligenza artificiale, su cui ha ripiegato anche il Consiglio d'Europa e, ancora, alla non più espressa previsione di una norma in materia di valutazione del rischio. Nel bene, se si considera invece lo spostamento dell'Unione Europea verso una impostazione più *human rights-based* rispetto alla versione originaria del testo pubblicata nell'aprile 2021.

In relazione a quest'ultimo aspetto, sembra che sia stato più il Consiglio d'Europa ad influenzare l'Unione Europea che non viceversa. Una influenza, però, che si è risolta in un più ampio spazio

³⁹ Così, l'art. 2.

⁴⁰ All'art. 2 si leggeva, infatti, quanto segue: «[i]n order to give full effect to the principles and obligations set out in this Convention, each Party shall maintain and take such graduated and differentiated measures in its domestic legal system as may be necessary and appropriate in view of the severity and probability of occurrence of adverse impacts on human rights and fundamental freedoms, democracy and the rule of law during design, development, use and decommissioning of artificial intelligence systems».

⁴¹ Per una analisi delle tendenze globali sulla *governance* dell'IA, si rinvia a H. ROBERTS, E. HINE, M. TADDEO, L. FLORIDI, *Global AI governance: barriers and pathways forward*, in *International Affairs*, 3, 2024, 1275 ss.

riservato a diritti ulteriori rispetto alla riservatezza e alla protezione dei dati – si pensi, tra tutti, al richiamo alla Carta dei diritti fondamentali dell’Unione nel Preambolo – e che, tuttavia, non è scevra di perplessità. La principale risiede nell’omesso riferimento alle Direttive in materia di diritto anti-discriminatorio dell’Unione⁴², che l’*AI Act* non menziona nemmeno nel Preambolo, lasciando aperto più di un interrogativo circa la riferibilità o meno di tale *corpus* normativo anche alla *AI-derived discrimination*. Qui, viceversa, la *Framework Convention* presenta una risposta più efficace, dedicando due norme ai principi di eguaglianza e di non discriminazione che agiscono secondo una prospettiva duplice: quale principio guida o generale a cui subordinare, da un lato, la creazione e l’impiego delle tecnologie di IA e, dall’altro, come diritto a cui guardare nella fase di implementazione dei sistemi in esame.

Sotto altra prospettiva, nonostante i testi abbraccino la stessa nozione di intelligenza artificiale, una differenza importante investe l’ambito applicativo della c.d. valutazione di impatto dei sistemi di intelligenza artificiale sui diritti fondamentali, che i due testi disciplinano, rispettivamente, ai sensi degli artt. 27, l’*AI Act*, e 16, la *Framework Convention*. Il legislatore dell’Unione ha, infatti, scelto di circoscrivere l’operato ai soli sistemi c.d. “ad alto rischio”⁴³, laddove, all’opposto, il Consiglio d’Europa non delimita la valutazione a tipologie più o meno “pericolose”, allargando così la valutazione a tutti i sistemi di IA e non introduce divieti circa l’impiego di talune tipologie di tecnologie.

La differenza, sebbene importante, va, però, inserita nel quadro della eterogenea vincolatività dei due strumenti.

In altre parole, la perplessità è che il carattere meno vincolante del trattato del Consiglio d’Europa rispetto al Regolamento dell’Unione Europea finisca con il rendere solo potenzialmente più protettivo il trattato, residuando ampio margine di apprezzamento agli Stati sulle modalità attraverso le quali darvi attuazione. L’art. 16 del trattato del Consiglio d’Europa, rischia, cioè, di rimanere una norma solo sulla carta più garantista, non aggiungendo in definitiva nulla a quanto dispone il vincolante a tutti gli effetti *AI Act*.

Da ultimo, così come la *Framework Convention* si astiene dal delimitare la valutazione di impatto ai sistemi “ad alto rischio”, allo stesso modo, il trattato omette di suggerire qualsiasi classificazione dei sistemi di intelligenza artificiale come, viceversa, propone, pur con tutti i suoi limiti, il Regolamento dell’Unione Europea. Ne discende, che il trattato non introduce deroghe al ricorso a sistemi di intelligenza artificiale ritenuti con maggiore frequenza causa di violazioni di diritti fondamentali – si pensi, per tutte, alle tecnologie di riconoscimento facciale –, anche se la previsione di cui all’art. 3, paragrafo

⁴² Il riferimento, in particolare, è alla c.d. *Race Directive*, Direttiva 2000/43/CE del Consiglio, del 29 giugno 2000, che attua il principio della parità di trattamento fra le persone indipendentemente dalla razza e dall’origine etnica, e alla Direttiva 2000/78/CE del Consiglio, del 27 novembre 2000, che stabilisce un quadro generale per la parità di trattamento in materia di occupazione e di condizioni di lavoro.

⁴³ Quelli, cioè, elencati all’Annex III del Regolamento e richiamati a norma dell’art. 6, § 2, dell’*AI Act*. In particolare, quanto ai sistemi di riconoscimento facciale, si rinvia a quanto precisa il *Recital* n. 54, dove si legge quanto segue: «[a]s biometric data constitutes a special category of personal data, it is appropriate to classify as high-risk several critical-use cases of biometric systems, insofar as their use is permitted under relevant Union and national law. Technical inaccuracies of AI systems intended for the remote biometric identification of natural persons can lead to biased results and entail discriminatory effects. The risk of such biased results and discriminatory effects is particularly relevant with regard to age, ethnicity, race, sex or disabilities. Remote biometric identification systems should therefore be classified as high-risk in view of the risks that they pose».

4, che esclude dall'ambito applicativo della *Framework Convention* ogni questione inerente la difesa nazionale, potrebbe essere letta in senso conforme all'eccezione introdotta dall'*AI Act* circa il legittimo impiego di sistemi di IA "ad alto rischio" in presenza di esigenze di sicurezza pubblica.

Vero è che questa norma del trattato pare criticabile al punto da minare in radice le potenzialità del trattato, comprimendone a monte la già precaria vincolatività a livello domestico.

In proposito, il precedente *Consolidated Working Draft* proponeva una formulazione preferibile. Il testo dell'allora art. 7, preposto all'elenco dei principi di cui all'attuale Capo III, possedeva una ratio differente, chiarendo meglio la finalità di eventuali e ammissibili eccezioni. Si stabiliva, cioè, che nessuna deroga ai principi salvaguardati dal trattato avrebbe potuto essere introdotta dagli Stati contraenti se non per ragioni, tra le altre, di sicurezza nazionale, difesa, sanità.

È vero che il Rapporto Esplicativo giustifica la deroga rifacendosi all'art. 1 dello Statuto del Consiglio d'Europa, precisando che la norma non esclude l'assoggettamento dei sistemi di IA al diritto internazionale. Tuttavia, tali affermazioni paiono in ogni caso contraddire ed indebolire la portata del trattato, legittimando scelte discrezionali, eterogenee e, eventualmente, meno garantiste di quegli stessi principi su cui si regge la Convenzione Quadro, minando il consolidamento di obiettivi comuni tra gli Stati Parte.

A voler tirare le fila di quanto appena descritto, sembra che a beneficiare degli scambi tra le due organizzazioni internazionali sia stato più l'*AI Act* rispetto al trattato del Consiglio d'Europa, che la versione approvata nel maggio 2024 ci restituisce debole, sicuramente meno pregnante e innovativa di quanto le versioni precedenti avessero fatto sperare. I due testi, cioè, si sono quasi scambiati le parti. Dalla *Zero Draft*⁴⁴, alla *Revised Zero Draft*⁴⁵ sino al *Consolidated Working Draft*, il Consiglio d'Europa sembrava, infatti, occupare una posizione di primo piano, almeno quanto alla esigenze di tutela dei diritti fondamentali e, invece, sul finire ha ceduto il passo all'Unione Europea senza che, purtroppo, nessuno dei due testi si dimostri attualmente pienamente soddisfacente. Di questo, lo scarso peso della non discriminazione nell'impianto dell'*AI Act* è solo un esempio, non però poco importante.

4. Riflessioni conclusive sull'impatto (davvero?) globale del Trattato, oltre il Consiglio d'Europa

Le settimane, che hanno preceduto l'entrata in vigore del trattato, sono state caratterizzate da alcuni eventi degni di nota.

Al di là dell'approvazione dell'*AI Act*, di cui si è già detto, merita ricordare che, sebbene optando per uno strumento di *soft law*, anche le Nazioni Unite hanno scelto di intervenire "dicendo la loro".

La prima risoluzione dell'ONU in materia di intelligenza artificiale è stata, così, pubblicata l'11 marzo 2024⁴⁶, poco dopo il Regolamento dell'Unione e due mesi prima l'adozione della Convenzione Quadro.

⁴⁴ Il testo della *Zero Draft* può essere consultato al seguente link: <https://www.statewatch.org/media/3697/coe-artificial-intelligence-convention-zero-draft-30-6-22.pdf> (ultima consultazione 29/11/2024).

⁴⁵ Su cui si veda il testo integrale al link: <https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f> (ultima consultazione 29/11/2024).

⁴⁶ Qui il link al testo della Risoluzione: <https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=ArxLTgaVGOUYe5P1e5&fe=true> (ultima consultazione 29/11/2024).

La Risoluzione non è, però, il solo evento da richiamare e che, in breve, testimonia la volontà di riportare i diritti al centro, legando – questa è sicuramente una novità –, per la prima volta, il tema dello sviluppo sostenibile all'intelligenza artificiale⁴⁷.

Maggiore peso riveste, però, nella prospettiva che qui interessa, la lettera aperta, sottoscritta da numerose associazioni non governative, che sollecitava il Consiglio d'Europa a ritornare sui suoi passi, emendando il testo del trattato, poi approvato, reo di essersi tramutato in un documento privo di contenuti e adeguatamente protettivo dei diritti esposti alle ricadute anche pregiudizievoli dei sistemi di intelligenza artificiale.

Il riferimento a questa lettera aperta supporta alcune riflessioni conclusive sull'esito delle negoziazioni e sulle aspettative, tradite, rispetto all'operato iniziale del Consiglio d'Europa, almeno sino all'adozione del *Consolidate Working Draft* di cui si è detto.

Con le parole “*No to the abdication of our rights*”, si apriva la lettera indirizzata dalle associazioni non governative firmatarie, nel gennaio del 2024, alle istituzioni dell'Unione Europea, nonché agli Stati membri parti attive della negoziazione affinché ostacolassero alcune delle criticità che il testo ha poi ratificato, anche sulla scorta delle pressioni, per tutti, degli Stati Uniti d'America, invitati a prendere parte al processo di definizione dei contenuti del trattato.

I due aspetti, su cui appuntavano le proprie preoccupazioni le associazioni non governative firmatarie dello *Statement*, sono condivisibili e vanno ripresi in chiusura a testimonianza della torsione del trattato e dell'indebolimento delle garanzie che le precedenti versioni del testo avevano, viceversa, tratteggiato in modo convincente.

Ci si riferisce, per prima, all'esenzione senza eccezioni dall'applicabilità del trattato e, quindi, alla completa esclusione dal suo ambito applicativo di tutte le ipotesi di impiego dei sistemi di intelligenza artificiale qualora venga impiegati per ragioni di difesa nazionale e di sicurezza. La deroga appare tanto più grave se si considera che i pochissimi casi sinora giunti con successo dinanzi alle Corti, nazionali e sovranazionali, hanno, non così casualmente, riguardato proprio l'utilizzo, talvolta massiccio, delle tecnologie di intelligenza artificiale per ragioni di sorveglianza pubblica, sfociando in condanne delle autorità di pubblica sicurezza per le conseguenze discriminatorie scaturenti da tali sistemi.

Non meno severe sono, in secondo luogo, le critiche relative all'altra pesante esenzione, che taglia fuori tutti gli enti privati, le *big tech* per prime, dal rispetto delle norme convenzionali. La scelta è evidentemente politica, ma non può certo posarsi simile retrocessione delle garanzie di salvaguardia dei diritti fondamentali di fronte ad esigenze di mercato.

Deludente, in definitiva, il trattato che ci consegna il Consiglio d'Europa.

Deludente quanto alle aspettative iniziali, e, soprattutto, se si pensa alle speranze che potesse sopprimere o, almeno, contenere alcune delle debolezze del coevo *AI Act*.

Si chiudono, allora, queste riflessioni con due passaggi dello *Statement* di cui sopra, che meglio di ogni altro esprimono le criticità della Convenzione Quadro:

⁴⁷ In questo senso, muove il § 3, punto *d*), dove così si legge: «[e]nhancing the ability of developing countries, in particular the least developed countries, to address major structural impediments and lift obstacles to accessing the benefits of new and emerging technologies and artificial intelligence innovation to achieve all 17 Sustainable Development Goals, including through scaling up the use of scientific sources, affordable technology, research and development, including through strengthened partnerships».

«[w]e have never given mandate to our elected representatives to abdicate our rights through the conclusion of a Convention that is ironically supposed to safeguard them. A hollowed-out Convention will provide little meaningful protections to individuals who are increasingly subject to powerful AI systems prone to bias, human manipulation, and the destabilisation of democratic institutions. As for public activities, some states are pushing for a blanket exemption with regard to national security and defence. Nothing justifies the unconditional waiving of the safeguards set in international, European and national law that usually apply in these fields. These attempts also turn a blind eye to the geopolitical context characterised by an increasing prevalence of dual use AI that put our lives and freedom at risk».

Se il futuro saprà riservare di meglio, è un augurio che, in questa sede, si ritiene di potere e dovere auspicare.

Special Issue



