

Lo studio di fattibilità di un nuovo quadro normativo sulla concezione, lo sviluppo e l'applicazione dei sistemi di Intelligenza Artificiale sulla base delle norme del Consiglio d'Europa

Il lavoro del Comitato Ad hoc sull'intelligenza artificiale del Consiglio d'Europa (CAHAI)

di Clementina Barbaro
Co-segretaria del CAHAI

Nel corso della sua terza riunione plenaria, il 15-17 Dicembre 2020, il *Comitato Ad hoc sull'intelligenza artificiale del Consiglio d'Europa (CAHAI)* ha adottato all'unanimità lo studio di fattibilità di un quadro normativo sulla concezione, lo sviluppo e l'applicazione dell'intelligenza artificiale (IA) fondato sulle norme del Consiglio d'Europa in materia di diritti dell'Uomo, democrazia e stato di diritto, così realizzando un compito primario del suo mandato¹. Lo studio di fattibilità espone le ragioni della necessità di un tale quadro normativo (I), iniziando al contempo una riflessione sui suoi elementi costitutivi (II), che il CAHAI si propone di affinare nel 2021, in vista della presentazione delle sue conclusioni al Comitato dei Ministri del Consiglio d'Europa. Tale istanza prenderà la decisione finale sullo strumento giuridico internazionale che sarà elaborato a partire dal 2022: tra le diverse opzioni possibili, quella di un trattato internazionale (III).

1. La necessità di un quadro normativo internazionale sull'IA fondato sulle norme del Consiglio d'Europa

Se lo studio omette di fornire una definizione precisa dell'IA, lasciandone l'opportunità ad uno stadio ulteriore dei lavori del CAHAI², rileva tuttavia il grande potenziale trasformativo delle tecnologie di IA sul piano individuale e collettivo, soffermandosi sui rischi e benefici potenziali con una serie di esempi concreti.

Ed è proprio la considerazione dei rischi incombenti sulla democrazia e lo stato di diritto (congiuntamente ad una definizione delle possibili risposte per mitigarli e prevenirli) a rappresentare l'originalità del lavoro del CAHAI rispetto a quelle di altre organizzazioni internazionali ed europee, come UE ed UNESCO, in cui l'accento è prevalentemente sull'impatto

¹ Il mandato del Comitato è «esaminare la fattibilità e gli elementi potenziali – sulla base di una larga consultazione multipartita – di un quadro normativo sulla concezione, lo sviluppo e l'applicazione dell'intelligenza artificiale (IA) fondato sulle norme del Consiglio d'Europa in materia di diritti dell'Uomo, democrazia e stato di diritto. Il CAHAI riunisce, oltre ai 47 Stati membri e 6 Stati osservatori, 72 rappresentanti di comitati ed istituzioni del Consiglio d'Europa, della società civile, del settore scientifico e delle compagnie facenti parte del Partenariato Digitale / Digital Partnership Framework del Consiglio d'Europa.

² Se lo studio del CAHAI fa riferimento ad alcune caratteristiche dei sistemi di IA (ad esempio la loro capacità di rilevare l'ambiente in cui operano attraverso l'acquisizione e l'analisi di dati, di elaborare delle informazioni da tali dati, nonché di decidere la migliore linea d'azione per raggiungere un certo obiettivo), gli esperti del CAHAI hanno convenuto di affrontare i sistemi di IA in modo tecnologicamente neutrale (cioè indipendentemente dalla tecnologia utilizzata), comprendendo tutte le varie tecnologie decisionali automatizzate considerate di IA e tenendo conto del loro contesto socio-tecnico di applicazione. Inoltre, il CAHAI ha rilevato la necessità di trovare un equilibrio tra una definizione troppo precisa da un punto di vista tecnico, che potrebbe quindi essere obsoleta a breve termine, e una definizione troppo vaga e che lascerebbe un ampio margine di interpretazione, potenzialmente risultante in un'applicazione non uniforme del quadro giuridico. Di conseguenza, un futuro quadro normativo del Consiglio d'Europa sull'IA dovrebbe adottare una definizione semplificata e tecnologicamente neutra che copra quelle pratiche o casi applicativi in cui lo sviluppo e l'uso di sistemi di IA, o di sistemi decisionali automatizzati più in generale, possono avere un impatto sui diritti umani, la democrazia e lo stato di diritto, e che tenga conto di tutte le implicazioni socio-tecniche di tali sistemi.

delle tecnologie di IA sui diritti fondamentali dell'individuo (unitamente ad altre considerazioni pertinenti al mandato e finalità di tali organizzazioni). Innegabilmente, l'analisi dell'incidenza di tali tecnologie sui diritti umani rappresenta un elemento centrale anche dello studio di fattibilità del CAHAI, in particolare per quanto attiene all'impatto negativo delle tecnologie di IA – di cui sono illustrati alcuni esempi - su una vasta gamma di diritti civili, politici, economici e sociali³: si pensi, ad esempio, al diritto alla libertà ed alla sicurezza, al diritto ad un equo processo e ad un rimedio effettivo; al diritto alla vita privata e familiare, all'integrità fisica, psicologica e morale; al diritto alla libertà di espressione ed alla libertà di associazione e di assemblea; l'uguaglianza e la non-discriminazione; il diritto al lavoro ed alle prestazioni sociali e mediche.

Lo studio sottolinea inoltre i rischi per il buon funzionamento dei processi democratici derivanti da alcune applicazioni di IA, che possono essere utilizzate per finalità di controllo dei cittadini – applicazioni di filtro e di classificazione di informazioni, ad esempio a fini di censura – o di sorveglianza di massa; sono rilevate anche certe criticità nell'ambito dei processi elettorali: se la propaganda e la manipolazione non sono di certo fenomeni nuovi, gli strumenti basati sull'IA permettono di amplificarle⁴ e di realizzare campagne di disinformazione sia ad ampia scala sia altamente personalizzate, in quanto fondate sulla profilazione di utenti spesso ignari del trattamento di dati contenenti informazioni dettagliate su interessi e preferenze personali⁵.

L'incidenza sullo stato di diritto è ugualmente oggetto di analisi: si rilevano in particolare i rischi per alcune garanzie fondamentali dell'equo processo (come l'indipendenza e l'imparzialità del giudice, il principio del contraddittorio, o il diritto all'assistenza di un avvocato – già rilevati nella Carta etica della CEPEJ sull'uso dell'intelligenza artificiale nei sistemi giudiziari ed ambiti connessi⁶, in particolare con riferimento ad applicazione cosiddette di «giustizia predittiva» in ambito civile e penale) che sono a loro volta pilastri dello stato di diritto.

A questa analisi di tipo sostanziale si aggiunge uno studio delle lacune normative che rendono necessario un nuovo quadro normativo sulla concezione, lo sviluppo e l'applicazione dell'IA fondato sulle norme del Consiglio d'Europa. Occorre rilevare *in primis* che non esiste, allo stato attuale, uno specifico strumento giuridico internazionale vincolante su questo tema, mentre esistono diversi strumenti giuridici internazionali - vincolanti e non vincolanti – applicabili d alcuni effetti dei sistemi di IA, come la Convenzione europea dei diritti dell'Uomo (CEDU), la Convenzione 108 del Consiglio d'Europa, la Carta sui diritti fondamentali dell'Unione europea.

Tuttavia, il numero e la diversità di questi strumenti rende complessa l'opera di interpretazione e di applicazione ai sistemi di IA. Inoltre, da un punto di vista sostanziale, i diritti e gli obblighi esistenti tendono ad essere articolati in modo molto ampio e generico, non garantendo così una risposta adeguata alle problematiche poste dai sistemi di IA. Si pone dunque un problema di contestualizzazione dei principi e dei diritti consacrati in norme spesso redatte anteriormente all'avvento dei sistemi di IA, già evocato da altre fonti e a cui lo studio cerca di dare una risposta esauriente nel suo capitolo 7.⁷

³ Si pensi ad esempio ad applicazioni fondate sul trattamento di dati personali o non personali in grado di profilare o categorizzare gli individui per vari scopi e in diversi contesti non compatibili con i diritti umani; per maggiori dettagli si veda il capitolo 3 dello studio di fattibilità, pag. 7-11.

⁴ Si veda a questo proposito la *Dichiarazione del Comitato dei Ministri sulla capacità manipolativa dei sistemi algoritmici*, «in grado non solo di prevedere le scelte, ma anche di influenzare emozioni e pensieri e alterare lo svolgimento di un'azione, a volte in modo subliminale»; *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, 13 February 2019.

⁵ Si veda il caso di Cambridge Analytica, che avrebbe utilizzato in modo illecito i dati di oltre 50 milioni di elettori americani profilandone psicologia e comportamento in base al monitoraggio delle loro attività su Facebook. The New York Times, *How Trump Consultants Exploited the Facebook Data of Millions* (consultato il 19/04/2021)

⁶ *Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari ed in ambiti connessi della European Commission for the Efficiency of Justice* – CEPEJ, Dicembre 2018.

⁷ Studio di fattibilità, capitolo 7. Si veda inoltre lo studio di Alessandro Mantelero incluso nella pubblicazione *Towards regulation of AI systems* edita dal CAHAI, Dicembre 2020, pag. 61 e seguenti. Inoltre, oltre alla

Inoltre, un certo numero di principi essenziali per la protezione dei diritti umani, della democrazia e dello stato di diritto nel contesto dell'IA non sono esplicitamente previsti dalle norme attuali. Essi includono per esempio la necessità di assicurare il controllo e la supervisione umana sulle applicazioni di IA, di garantire la loro robustezza tecnica, e di assicurare la loro effettiva trasparenza e spiegabilità⁸. Tali lacune hanno a loro volta un'incidenza sui diritti fondamentali, la cui effettività rischia di essere compromessa qualora le decisioni prese dai sistemi di IA non offrano garanzie adeguate di trasparenza o di spiegabilità.

Queste lacune dimostrano che un quadro di *governance* dei sistemi di IA è di fondamentale importanza. La questione chiave non è la tecnologia di per sé, ma il suo utilizzo senza che siano fornite ai progettisti di sistemi IA delle regole chiare di sviluppo conformi ai diritti umani, alla democrazia e allo stato di diritto. A questo proposito, lo studio rileva che le lacune precedentemente menzionate generano incertezza proprio per tali soggetti, che non dispongono di un quadro normativo prevedibile e solido di progettazione e di attuazione di sistemi di IA; tale incertezza rischia di ostacolare una innovazione conforme agli standard del Consiglio d'Europa, socialmente benefica dei sistemi di IA e percepita dai cittadini come degna di fiducia⁹.

Se le linee guida etiche e gli strumenti di *soft law* elaborati nel corso degli ultimi anni - 323 secondo l'ultimo rendiconto fornito dal Consiglio d'Europa¹⁰ - hanno contribuito a strutturare il dibattito sui principi fondamentali che dovrebbero guidare la concezione e lo sviluppo delle applicazioni di IA, il carattere non vincolante di tali linee guida e strumenti, spesso privi di meccanismi di attuazione e di controllo adeguato, non possono essere considerati sufficienti. Peraltro, la formulazione di tali principi varia in modo considerevole da un testo all'altro: ad esempio, il principio di trasparenza (il principio più frequentemente enunciato secondo uno studio commissionato dal CAHAI¹¹) può essere inteso, a seconda del testo di riferimento, come trasparenza in senso stretto (accesso ai dati e al sistema algoritmico), come spiegabilità o interpretabilità del sistema di IA, o come comunicazione di informazione relativa al sistema di IA. Le stesse difficoltà si pongono con riferimento al principio di «giustizia ed equità» delle applicazioni di IA, che ingloba concetti molto diversi tra loro quali l'uguaglianza, la non-discriminazione, la diversità, l'accessibilità ed il diritto ad un rimedio effettivo.

Si evince quindi l'importanza di una risposta sul piano normativo che permetta di stabilire, in modo univoco, il contenuto di principi specificamente applicabili alla concezione, allo sviluppo e all'implementazione dei sistemi di IA, comuni a diverse aree e trasversali, fonti di diritti ed obblighi per gli Stati, che possano a loro volta servire di riferimento ad interventi settoriali specifici.

Secondo lo studio di fattibilità, una risposta di *governance* effettiva deve essere fondata su un insieme di strumenti giuridici che si completino e rinforzino reciprocamente, di natura vincolante e non vincolante. Uno strumento vincolante di carattere

sopracitata Carta etica della CEPEJ, si veda anche lo studio pubblicato dal CCBE *on the Legal Aspects of AI* (February 2020):

https://www.ccbe.eu/fileadmin/speciality_distribution/public/documents/IT_LAW/ITL_Guides_recommendations/EN_ITL_20200220_CCBE-considerations-on-the-Legal-Aspects-of-AI.pdf: a titolo di esempio, il principio del contraddittorio, laddove siano utilizzati dei sistemi di IA nell'ambito di un procedimento giudiziario, dovrebbe potere tradursi nella possibilità di contestare e dibattere le conclusioni prodotte dall'applicazione di IA prima della deliberazione del giudice.

⁸European Commission White Paper on AI, 19 February 2020, COM(2020) 65 final, at p 9: "A key result of the feedback process is that, while a number of the requirements are already reflected in existing legal or regulatory regimes, those regarding transparency, traceability and human oversight are not specifically covered under current legislation in many economic sectors".

⁹ Significativo a questo riguardo il risultato del processo di consultazione sul *White Paper* pubblicato dalla Commissione Europea in febbraio 2020 (<https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence>) : il rapporto finale sottolinea che solo il 2,9% degli intervistati considera l'attuale normativa dell'Unione Europea in grado di rispondere alle preoccupazioni generate dai sistemi IA.

¹⁰ *AI and digital policies framework*, consultato il 19/04.2021.

¹¹ M. Ienca, E. Vayena, *AI Ethics Guidelines: European and Global Perspectives*, studio incluso nella pubblicazione *Towards Regulation of AI systems*, p. 38 et ss.

orizzontale, come una Convenzione o una Convenzione Quadro ¹² potrebbe consolidare i principi comuni generali, mentre ulteriori strumenti settoriali vincolanti o non vincolanti potrebbero affrontare le sfide poste dai sistemi di IA in settori specifici, come ad esempio la giustizia.

Questo approccio è stato sostenuto anche [dall'Assemblea parlamentare del Consiglio d'Europa](#), che ha invitato il Comitato dei Ministri a sostenere la stesura di uno strumento giuridicamente vincolante che regoli le applicazioni di IA, possibilmente nella forma di una Convenzione del Consiglio d'Europa. Si tratta di una posizione politicamente importante, considerando il ruolo centrale che avranno i parlamenti nazionali nelle procedure di ratifica di tale strumento. Inoltre, sostegno per tale soluzione si trova anche nelle conclusioni della recente conferenza di alto livello su "Diritti umani nell'era dell'IA - L'Europa come promotrice di standard internazionali per l'intelligenza artificiale" organizzata dalla presidenza tedesca del Comitato dei Ministri del Consiglio d'Europa.

Se la certezza e la coerenza per gli operatori di IA sono obiettivi importanti, l'adozione di un approccio proporzionale, che si concentri sulle applicazioni cosiddette «ad alto rischio» per i diritti umani, la democrazia e lo stato di diritto, ha registrato il consenso dei membri del CAHAI.

2. Gli elementi costitutivi del futuro quadro giuridico.

Lo studio di fattibilità ha identificato alcuni principi chiave¹³ applicabili ai sistemi di IA che potrebbero essere considerati come elementi costitutivi di un futuro quadro normativo. Da questi principi derivano una serie di diritti ed obblighi per gli Stati (alcuni già consacrati dai testi esistenti, altri di nuova concezione e contestualizzati alle sfide ed opportunità posti dai sistemi di IA).

Il primo principio è quello della **dignità umana**, uno dei pilastri dell'architettura della CEDU: la sua declinazione rispetto ai sistemi di IA implica il diritto di essere informati dell'attivazione di un'interazione in corso con un sistema di IA e non con un essere umano – e della possibilità di rifiutarla - in particolare quando ci sia un rischio di confusione e di incidenza negativa sulla dignità umana. A questo diritto si contrappone l'obbligo di informazione, per gli sviluppatori, di tale interazione qualora vi sia possibilità di confusione sulla natura dell'interlocutore. Inoltre, gli Stati dovrebbero garantire che certe attività ritenute a rischio di violazione della dignità umana siano riservati agli esseri umani, e non ai sistemi di IA (si pensi ad esempio a compiti di assistenza di persone anziane o disabili).

La **prevenzione del danno** ai diritti umani, alla democrazia e allo stato di diritto, è un altro principio fondamentale. Gli Stati membri dovrebbero garantire che siano adottate – da parte degli sviluppatori di sistemi di IA - adeguate salvaguardie per ridurre al minimo e prevenire i danni derivanti dallo sviluppo e dall'uso dell'IA, che si tratti di danni fisici, psicologici, economici, ambientali, o sociali. Per questo, dovrebbero essere garantiti dagli sviluppatori e dai distributori di sistemi di IA adeguati requisiti di sicurezza, protezione e robustezza (*by design*) e il loro rispetto attraverso procedure di test e di verifica prima dell'utilizzo e durante l'intero ciclo di vita del sistema di IA; e particolarmente nel contesto delle procedure di appalto pubblico.

Nell'attuare le misure per prevenire i danni, gli Stati membri dovrebbero perseguire un approccio basato sul rischio, concentrandosi in particolare sulle applicazioni cosiddette “ad alto rischio” di violazione dei diritti umani, della democrazia e dello stato di diritto.

Inoltre, si rileva l'esigenza di garantire la sostenibilità¹⁴ dei sistemi di IA, tanto in fase di sviluppo

¹² Sulle differenze tra Convenzione e Convenzione Quadro si veda ulteriormente nella sezione III.

¹³ Per un'esposizione completa di tali principi si rinvia al Capitolo 7 dello studio; la descrizione dei principi nell'ambito di questa sessione rappresenta in effetti una sintesi di alcuni punti salienti dello studio.

¹⁴ L'addestramento dell'intelligenza artificiale è un processo ad alta intensità energetica. Nuove stime suggeriscono che l'impronta di carbonio dell'addestramento di una singola IA è di ben 284 tonnellate di

che di applicazione.

La libertà e l'autonomia umana nel contesto dell'IA debbono essere intese come la capacità degli esseri umani di agire e di decidere in modo informato e autonomo sull'uso dei sistemi di IA e sulle loro conseguenze. Questo include anche le decisioni sul se, quando e come usare i sistemi di IA, nonché di garantire adeguati meccanismi di supervisione umana. Tra i diritti sostanziali enunciati figurano il diritto di decidere liberamente di essere esclusi dalla manipolazione, dalla profilazione individualizzata e dalle previsioni dell'IA, anche in caso di trattamento di dati non personali; nonché il diritto di avere l'opportunità, qualora non sia escluso da motivi imperativi legittimi concorrenti, di scegliere di avere contatti con un essere umano piuttosto che con un robot.

Lo studio inoltre elenca una serie di misura volte ad assicurare la **non-discriminazione, l'uguaglianza di genere, l'equità e la diversità** dei sistemi di IA. Si sofferma in particolare sul rischio di discriminazione intersezionale derivante dall'uso di tali sistemi¹⁵ e sottolinea l'opportunità di includere tale discriminazione tra gli elementi del diritto alla non-discriminazione, in aggiunta ai motivi di cui all'articolo 14 della CEDU e al protocollo 12 della CEDU. Rileva inoltre un'ulteriore criticità (ed una sfida sul piano normativo) legata alla capacità dei sistemi di IA di generare un trattamento ingiusto basato su nuovi tipi di differenziazione che non sono tradizionalmente protetti¹⁶.

In ogni caso, tenendo conto anche di numerosi esempi di applicazione¹⁷, lo studio evidenzia che il principio di non-discriminazione deve essere garantito in relazione all'intero ciclo di vita di un sistema di IA (progettazione, sviluppo, implementazione e utilizzo). Gli Stati dovrebbero applicare il più alto livello di scrutinio qualora utilizzino o promuovano l'uso di sistemi di IA in settori sensibili di politica pubblica, per esempio la giustizia, l'asilo e la migrazione, la salute, la sicurezza sociale e l'occupazione, introducendo salvaguardie e requisiti specifici al fine di prevenire possibili effetti

biossido di carbonio equivalente a cinque volte le emissioni di una autovettura di medio calibro. Per approfondimenti, si veda lo studio di University of Massachusetts Amherst, USA, <https://www.newscientist.com/article/2205779-creating-an-ai-can-be-five-times-worse-for-the-planet-than-a-car/#ixzz6pBsHZ9JG>

¹⁵ La discriminazione intersezionale avviene sulla base di diversi motivi o caratteristiche personali che operano e interagiscono tra loro allo stesso tempo in modo tale da essere inseparabili. Gli attuali sistemi di intelligenza artificiale sono particolarmente suscettibili a tale discriminazione, in quanto cercano semplicemente delle correlazioni tra diverse caratteristiche. Il futuro quadro normativo del Consiglio d'Europa dovrebbe rivolgere un'attenzione particolare a questa questione, poiché la discriminazione intersezionale è raramente coperta dalle leggi nazionali sulla discriminazione, che tendono a concentrarsi su ragioni di discriminazioni trattate singolarmente (ad esempio sesso, razza, colore, lingua, religione, etc.).

¹⁶ Si veda l'esempio nello studio del CoE su AI e discriminazione studio di F. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, 2018, a p.35: «Supponiamo che un sistema di AI trovi una correlazione tra (i) l'uso di un certo browser web e (ii) una maggiore disponibilità a pagare. Un negozio online potrebbe applicare prezzi più alti alle persone che usano quel browser. Tali pratiche rimangono al di fuori del campo di applicazione della legge sulla non discriminazione, in quanto il tipo di browser non è una caratteristica protetta».

¹⁷ Nell'ambito dei sistemi giudiziari, sono stati particolarmente dibattuti gli esempi di polizia predittiva e di strumenti di localizzazione criminale predittiva, come il programma *PredPol*, uno dei primi operanti ma successivamente abbandonato, ufficialmente per ragioni legate alla pandemia di COVID-19, ma criticato in considerazione degli effetti discriminatori su alcune comunità (latino-americane e afro-americane: <https://www.latimes.com/california/story/2020-04-21/lapd-ends-predictive-policing-program>) e delle cosiddette "profezie autorealizzatrici": i quartieri considerati a rischio attirano maggiormente l'attenzione della polizia, che scopre conseguentemente un maggior numero di reati, con il risultato di un'eccessiva sorveglianza da parte della polizia delle comunità residenti in tali luoghi. Si vedano inoltre gli esempi degli strumenti di predizione del rischio di recidiva: gli esperimenti effettuati in altri Paesi (COMPAS negli Stati Uniti e HART nel Regno Unito) sono stati criticati da alcune ONG (si vedano i lavori di ProPublica negli Stati Uniti e di Big Brother Watch nel Regno Unito) a causa dei limiti delle metodologie utilizzate, che hanno prodotto risultati errati: la constatazione del fatto che alcuni individui afroamericani sono coinvolti più spesso in atti criminali ha comportato un maggiore fattore di rischio per l'intera popolazione afroamericana. Pertanto, anche in relazione a reati minori, tali sistemi hanno valutato negativamente gli imputati afroamericani, e ciò ha dato luogo a un iniquo aumento della durata delle pene inflitte a essi.

discriminatori: ad esempio, la trattazione delle lacune esistenti nei dati di genere, la rappresentatività, la qualità e l'accuratezza dei set di dati, la funzione di progettazione e ottimizzazione degli algoritmi, il monitoraggio dell'uso del sistema, e adeguati processi di test e valutazione per verificare e mitigare i rischi di discriminazione.

Per quanto riguarda **la trasparenza e la spiegabilità** dei sistemi di IA, lo studio sottolinea l'importanza di creare diritti individuali specifici: un diritto all'informazione qualora una decisione che produce effetti giuridici o ha un impatto significativo sulla vita di un individuo sia stata informata o presa da un sistema di IA, a cui si unisce il diritto ad una spiegazione significativa di come funzioni tale sistema, quale logica di ottimizzazione segua, che tipo di dati utilizzi e come influisca sugli interessi individuali.

A questi diritti si contrappongono alcuni obblighi per gli Stati e per gli sviluppatori dei sistemi di IA per quanto riguarda la tracciabilità e la fornitura di informazioni, che dovrebbero essere comprensibili e accessibili e potrebbero, tra l'altro, includere i tipi di decisione o situazione soggetti al trattamento automatizzato, i criteri rilevanti per una decisione, le informazioni sui dati utilizzati, una descrizione del metodo di raccolta dei dati. Una descrizione dei potenziali effetti dell'applicazione dovrebbe essere accessibile per la revisione/audit da parte di organismi indipendenti con le necessarie competenze. Per quanto riguarda la documentazione, le serie di dati e i processi che producono le decisioni del sistema di IA (raccolta ed etichettatura dei dati, algoritmi utilizzati), dovrebbero essere documentati, consentendo così la verificabilità *ex post* del sistema. Si tratta di requisiti particolarmente importanti qualora i sistemi di IA siano utilizzati nella fornitura di servizi pubblici, e lo studio sottolinea la necessità di trovare un bilanciamento tra i segreti commerciali e i diritti di proprietà intellettuale ed altri interessi legittimi. A questo proposito, le autorità pubbliche devono essere in grado di controllare i sistemi di IA qualora vi siano solide indicazioni di non conformità, per verificare il rispetto della legislazione esistente. Tuttavia, si rileva che gli oneri tecnici di trasparenza e spiegabilità non dovrebbero limitare in modo irragionevole le opportunità di mercato, laddove i rischi per i diritti umani, la democrazia e lo stato di diritto siano meno importanti.

La protezione dei dati ed il diritto alla privacy possono essere messe a repentaglio dai sistemi di IA, anche qualora tali sistemi di IA non siano progettati per elaborare dati personali e si basino invece su dati anonimizzati, anonimi o non personali: in effetti, i sistemi di apprendimento automatico possono dedurre da tali dati informazioni personali, compresi i dati sensibili. Gli Stati membri devono dunque garantire che il diritto alla privacy e alla protezione dei dati sia salvaguardato durante l'intero ciclo di vita dei sistemi di IA: a questo proposito si ricorda che i principi stabiliti dalla Convenzione 108+ (tra cui equità e trasparenza, proporzionalità, legittimità del trattamento, qualità dei dati, diritto a non essere soggetti a decisioni puramente automatizzate e altri diritti della persona interessata, sicurezza dei dati, responsabilità, valutazioni d'impatto e privacy by design) sono applicabili al trattamento dei dati personali in qualsiasi fase del ciclo di vita di un sistema di IA, compresi i dati.

Una particolare attenzione è rivolta ai rischi di sorveglianza di massa derivanti dai sistemi di IA, ad esempio attraverso la tecnologia di riconoscimento biometrico a distanza o altre tecnologie di tracciamento: su questo tema, ha preso posizione recentemente proprio il Comitato della Convenzione 108 con l'adozione di Linee Guida specifiche¹⁸.

¹⁸ Si vedano le [Guidelines on Facial Recognition](#), 28 Gennaio 2021, che rilevando i rischi significativi per la privacy e la protezione dei dati posti dall'uso crescente di queste tecnologie, prevedono di inquadrare strettamente e eventualmente di proibire certi usi, come l'uso diretto – e non accompagnato da garanzie adeguate previste dalla legge - esclusivamente a determinare il colore della pelle, le credenze religiose o di altro tipo, il sesso, l'origine razziale o etnica, l'età, la salute o lo stato sociale di una persona.

Questo divieto dovrebbe ugualmente applicarsi alle tecnologie di "riconoscimento degli affetti", che identificano le emozioni e possono essere utilizzate per rilevare i tratti della personalità, i sentimenti più intimi, la salute mentale o l'impegno dei dipendenti, in quanto rappresentano rischi significativi in settori come l'occupazione, o l'accesso all'assicurazione o all'istruzione.

Il principio di **responsabilità** implica che gli stati membri debbano adottare delle misure volte a garantire che i sistemi di IA, sin dalla fase della loro concezione che nella fase di diffusione e di utilizzo, siano conformi alle norme del Consiglio d'Europa. Nella stessa misura, dei meccanismi di responsabilità civile e penale per le eventuali violazioni causate dagli sviluppatori o dai distributori di sistemi di IA dovrebbero essere previsti, così come dei rimedi efficaci ed accessibili, che dovrebbero comportare la riparazione di qualsiasi danno subito, potendo includere misure di diritto civile (come l'obbligo di porre fine alla condotta illegale e le garanzie di non ripetizione, così come l'obbligo di risarcire il danno causato), amministrativo o, se del caso, penale.

Per quanto riguarda la **democrazia**, sono proposti una serie di obblighi per gli stati volti a trattare i rischi di interferenza nei processi elettorali, di manipolazione dell'opinione pubblica, di disinformazione e di discorsi d'odio online. Inoltre, si sottolinea l'importanza di definire parametri per l'azione pubblica che utilizzi sistemi di IA e favorire la partecipazione dei cittadini nei relativi processi decisionali. Particolarmente rilevanti sono le considerazioni in materia di appalti pubblici, che dovrebbero essere guidati da principi di trasparenza, di equità, di responsabilità e di affidabilità. Qualora il settore pubblico utilizzi i sistemi di IA, è necessario mobilitare competenze adeguate e interdisciplinari, per garantire una comprensione approfondita delle potenziali implicazioni dell'uso di tali sistemi per il buon governo e il rapporto tra il cittadino e la pubblica amministrazione. Inoltre, gli Stati dovrebbero rendere pubbliche e accessibili tutte le informazioni pertinenti sui sistemi di IA (compreso il loro funzionamento, il funzionamento di ottimizzazione, la logica sottostante, il tipo di dati utilizzati) che sono utilizzati nella fornitura di servizi pubblici, in un giusto bilanciamento con interessi legittimi come la sicurezza pubblica. Infine, gli Stati dovrebbero sottoporre l'uso dei sistemi di IA nel settore pubblico a meccanismi di controllo e di supervisione da parte di istanze giudiziarie o extra-giudiziarie.

Garanzie ed obblighi sono previsti anche con riferimento **alla giustizia ed i sistemi giudiziari**, i cui valori e principi fondamentali (indipendenza ed imparzialità del giudice, equo processo e contraddittorio) devono essere preservati qualora si faccia ricorso ad applicazioni di IA nel processo decisionale. A tal fine, si deve tenere conto della necessità di garantire la qualità e la sicurezza delle decisioni giudiziarie e dei dati che fanno l'oggetto di un trattamento automatizzato, così come la trasparenza, l'imparzialità e l'equità dei metodi di trattamento. A tal fine, dovrebbero essere introdotte garanzie per l'accessibilità e la spiegabilità dei metodi di trattamento dei dati, compresa la possibilità di audit esterni.

Qualora i sistemi di IA siano utilizzati in ambito extra-giudiziario a fondamento di un processo decisionale, l'accesso al giudice dovrebbe essere garantito. Qualora invece i sistemi di IA siano utilizzati in ambito giudiziario come supporto alla decisione del giudice, l'accesso delle parti alle informazioni pertinenti è di grande importanza. Tale accesso può riguardare, se necessario, i dati di addestramento del sistema di IA ed informazioni significative e comprensibili su come il sistema di IA ha raggiunto una raccomandazione, decisione o previsione. Dovrebbero inoltre essere fornite informazioni dettagliate sull'eventuale utilizzo di tali elementi nel processo decisionale del giudice.

3. Il lavoro del CAHAI nel 2021-2022

Se per il CAHAI il 2020 è stato l'anno della conferma della necessità di un quadro normativo sull'IA, il 2021 sarà l'anno dell'elaborazione degli elementi principali di tale quadro normativo, sulla base dei principi precedentemente menzionati ma anche delle indicazioni che saranno date dagli Stati membri. Un Gruppo di lavoro sul Quadro normativo (CAHAI-LFG¹⁹) si è riunito il 10-11 Febbraio 2021 per iniziare il lavoro di redazione di tali elementi²⁰, che saranno trasmessi al Comitato dei

Le linee guida sottolineano la necessità di un dibattito democratico sull'uso del riconoscimento facciale al volo nei luoghi pubblici e nelle scuole, data la natura intrusiva di questa tecnologia ed eventualmente la necessità di una moratoria in attesa di ulteriori analisi.

¹⁹ Legal Framework Group, per maggiori informazioni consultare il sito del CAHAI: www.coe.int/cahai

²⁰ La Professoressa Peggy Valcke, co-Presidente del CAHAI (Belgio) ed l Consigliere di Legazione Riccardo

Ministri del Consiglio d'Europa al termine del mandato del CAHAI (attualmente previsto per Dicembre 2021).

Oltre al CAHAI-LFG, il Gruppo di lavoro sull'Elaborazione delle Politiche (CAHAI-PDG²¹) si occuperà nel 2021 di questioni importanti come l'elaborazione di un modello di valutazione dell'impatto dei sistemi di IA sui diritti umani, la democrazia e lo stato di diritto, nonché di politiche relative all'uso dei sistemi di IA nella pubblica amministrazione.

Inoltre, una consultazione ad ampio raggio di diversi attori - tra cui la società civile e il settore privato - sul contenuto del quadro normativo è in corso e farà l'oggetto di uno specifico rapporto al CAHAI²².

I cantieri di lavoro aperti sono dunque diversi, e il lavoro del CAHAI proseguirà in coordinamento stretto con i membri del CAHAI e le altre Organizzazioni internazionali attive sul tema dell'IA, quali l'Unione Europea, l'OCSE e le Nazioni Unite.

Inoltre, una possibile evoluzione del mandato del CAHAI sarà oggetto di discussione nei prossimi mesi. Quando il Comitato avrà trasmesso gli elementi principali sopracitati al Comitato dei Ministri, questa istanza deciderà se autorizzare l'inizio del lavoro di stesura e di negoziazione di una futura Convenzione internazionale, o un altro tipo di strumento giuridico. Questa decisione sarà, molto probabilmente, presa nell'ambito della Presidenza italiana del Comitato dei Ministri del Consiglio d'Europa (Novembre 2021- Maggio 2022). Il Comitato dei Ministri potrà considerare diverse opzioni, anche se lo studio di fattibilità ha chiaramente sottolineato il valore aggiunto di una Convenzione o di una Convenzione Quadro a cui si abbinerebbero specifici strumenti settoriali vincolanti e non-vincolanti.

Sia le Convenzioni che le Convenzioni Quadro sono trattati multilaterali, hanno la stessa natura giuridica e sono entrambi soggetti alle regole relative ai trattati internazionali stabilite dalla Convenzione di Vienna sul diritto dei trattati (1969). Inoltre, entrambi possono includere un sistema di *governance* e possono essere integrati da protocolli aggiuntivi. La differenza è che una Convenzione regola una specifica materia o area in modo più concreto, tipicamente creando diritti e obblighi, mentre una Convenzione Quadro stabilisce piuttosto principi più ampi e aree di azione che sono state concordate tra gli Stati partecipanti. Una Convenzione Quadro tipicamente prevede solo un obbligo generale per gli Stati parte di intraprendere certe azioni, raggiungere certi obiettivi o riconoscere certi diritti, senza attribuire tali diritti direttamente alle persone fisiche o giuridiche. C'è quindi un notevole margine di discrezione per gli Stati su come attuare i principi e gli obiettivi della Convenzione Quadro.

Il Consiglio d'Europa ha dimostrato in passato di essere un'organizzazione pionieristica nel campo della bioetica, della protezione dei dati e della criminalità informatica, e attraverso il lavoro del CAHAI, desidera affrontare le sfide di *governance* poste dai sistemi di IA, che, come rilevato dall'Avvocato Paola Severino, già Ministro della Giustizia, «sprigioneranno infatti problemi tecnico-scientifici, interrogativi filosofico-normativi, considerazioni geo-politiche, profili legati all'etica dell'uomo nel rapporto con la macchina. Occorre allora cogliere tutti i prismatici riflessi di un fenomeno multidisciplinare per utilizzare le opportunità di uno sviluppo ormai inarrestabile mitigandone le conseguenze negative attraverso una adeguata regolamentazione. Se non sapremo raccogliere questa sfida ed affrontare fino in fondo le implicazioni etiche, economiche, giuridiche, che essa comporta, basandone la soluzione su una piena conoscenza delle tecnologie connesse all'uso della IA, potremmo trasformare un enorme potenziale di crescita in una discesa verso quel buio delle coscienze, delle democrazie e dei valori, così ben descritto nei libri di Noah Harari come l'incubo del XXI secolo²³».

Villa del Ministero degli Affari Esteri (Italia) sono stati eletti co-Presidenti.

²¹ Policy Development Group, per maggiori informazioni consultare il sito del CAHAI: www.coe.int/cahai

²² Si prega di consultare il sito del CAHAI (www.coe.int/cahai) per ulteriori informazioni.

²³ Paola Severino, Etica e macchine, 11 Aprile 2021, La Stampa <https://www.lastampa.it/topnews/lettere-e-idee/2021/04/11/news/etica-e-macchine-1.40135585>

