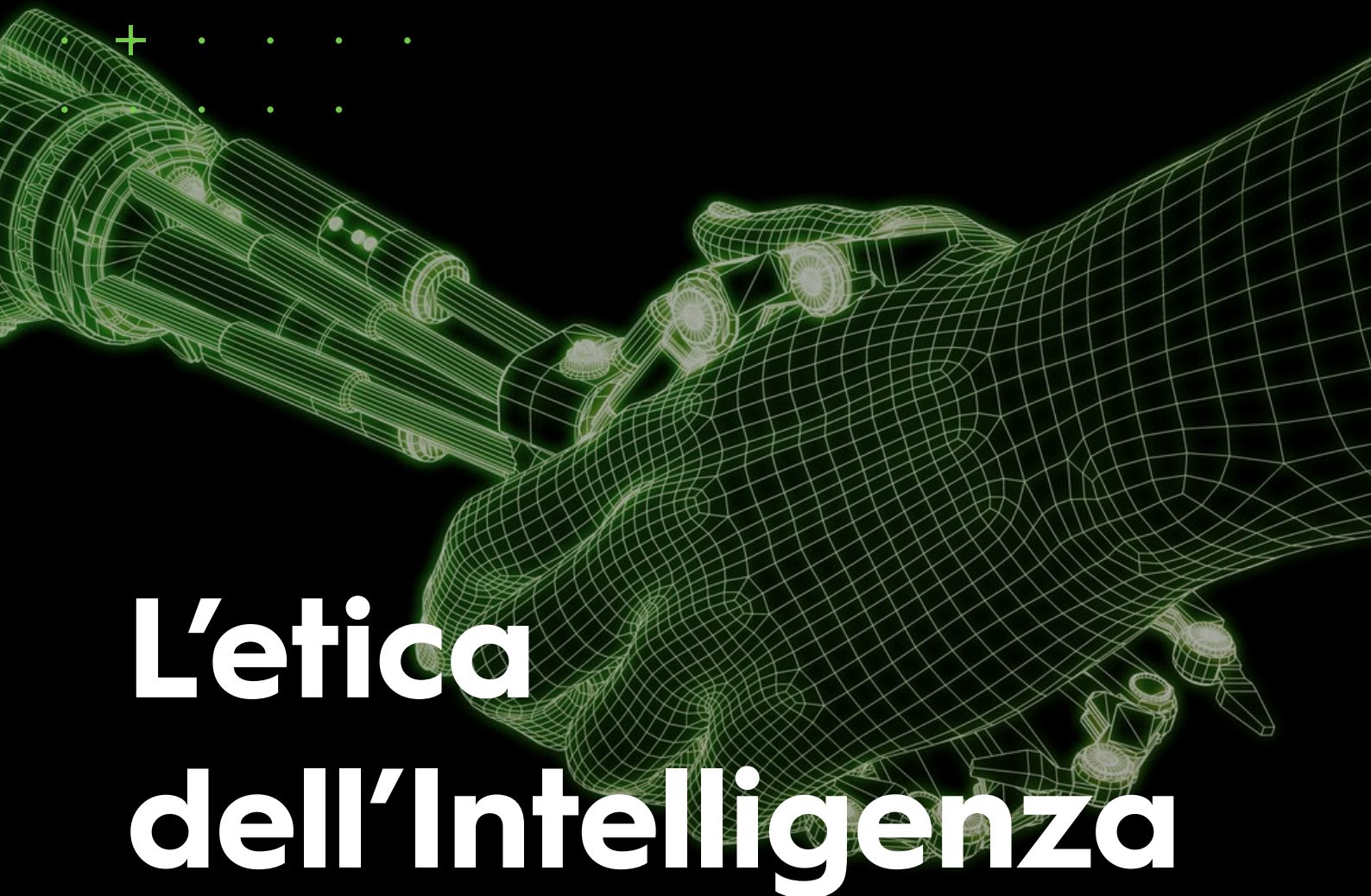


I QUADERNI DI **TECH4FUTURE**



L'etica dell'Intelligenza Artificiale

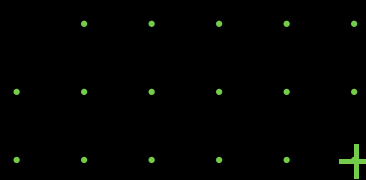
in questo numero

La Raccomandazione
dell'UNESCO
del 25 novembre
2021

I livelli di
rischio e il
primo quadro
giuridico dell'UE

La strategia
italiana
per un'IA
antropocentrica

Riflessioni attorno
alle conseguenze
eticamente
problematiche dell'IA



+ sommario

00

Prefazione
di Nicoletta Boldrini [leggi >](#)

01

La Raccomandazione dell'UNESCO sull'etica
dell'Intelligenza Artificiale [leggi >](#)

02

L'Europa definisce i livelli di rischio dell'Intelligenza
Artificiale e propone il primo quadro giuridico sull'IA [leggi >](#)

03

La strategia italiana per un'IA antropocentrica,
affidabile e sostenibile [leggi >](#)

04

Lo "Statuto etico e giuridico dell'Intelligenza Artificiale"
di Fondazione Leonardo-Civiltà delle Macchine [leggi >](#)

05

È etica l'Intelligenza Artificiale che fa del bene
all'intero ecosistema [leggi >](#)

06

Riflessioni attorno alle conseguenze eticamente
problematiche dell'IA [leggi >](#)

07

È il momento di passare dai principi alle
azioni concrete [leggi >](#)

08

Conclusioni
di Paola Cozzi [leggi >](#)

Prefazione

di Nicoletta Boldrini

Il primo dei Quaderni di Tech4Future vede una raccolta di articoli in tema di etica e Intelligenza Artificiale - pubblicati sulla nostra testata giornalistica nel corso dell'ultimo anno e mezzo - che abbiamo voluto riproporre inserendoli in un tessuto espositivo coerente, quasi un flusso narrativo che, da raccomandazioni, linee guida e strategie definite da Organismi internazionali e nazionali, approda a considerazioni che tengono conto del singolare momento storico in cui viviamo e a riflessioni di carattere filosofico.

Il dibattito sull'etica dell'Intelligenza Artificiale è, in realtà, un lavoro in corso. Iniziato in sordina una decina di anni fa, con lo sviluppo e l'immissione sul mercato di sempre nuovi algoritmi, destinati all'applicazione in un numero, via via, più ampio di settori, non si è più arrestato. Né mai si arresterà. È ormai parte del paradigma dell'Intelligenza Artificiale, della sua dialettica.

Dai primi dubbi sui risvolti poco nitidi, sotto il profilo etico, delle tecniche AI, al primo testo mondiale accettato da tutti gli Stati membri dell'ONU, che fissa standard condivisi volti a proteggere e a promuovere i diritti umani di fronte all'utilizzo di sistemi che sfruttano le tecniche di Intelligenza Artificiale, di strada ne è stata fatta. Il nostro compito è quello di raccontarne i percorsi, captando contenuti e accenti nuovi dovuti al mutare dei tempi (ad esempio, l'emergenza pandemica - come vedremo - ha modificato la sensibilità del dibattito), all'evoluzione delle tecnologie e all'integrazione con altre discipline e altri saperi.

Qui tracciamo un pezzo del percorso di tale "strada", con l'obiettivo non solo di fare informazione (è il nostro mestiere), bensì anche di suscitare in chi leggerà nuove domande sul lato chiaro e sul lato oscuro delle tecniche che fanno capo all'ambito di studi dell'Intelligenza Artificiale.

Grazie per il tuo tempo. Buona lettura.

La Raccomandazione dell'UNESCO sull'etica dell'Intelligenza Artificiale

La Conferenza Generale dell'Organizzazione delle Nazioni Unite per l'Educazione, la Scienza e la Cultura (UNESCO) - riconoscendo i profondi e dinamici impatti positivi e negativi dell'Intelligenza Artificiale (IA) sulle società, l'ambiente, gli ecosistemi e le vite umane, compresa la mente umana, in parte a causa dei nuovi modi in cui il suo uso influenza il pensiero umano, l'interazione e il processo decisionale e colpisce l'istruzione, le scienze umane, sociali e naturali, la cultura, la comunicazione e l'informazione - ha adottato un documento di Raccomandazione sull'etica dell'Intelligenza Artificiale.

Come tutti i documenti UNESCO e le Raccomandazioni dell'ONU in generale, anche l'applicazione di questa carta è su base volontaria. Tuttavia, **la Raccomandazione rilasciata il 25 novembre 2021**, primo testo mondiale accettato e ratificato globalmente da tutti i 193 Stati membri dell'Organizzazione delle Nazioni Unite, **rappresenta un tassello importante per la definizione di standard condivisi, volti a proteggere e a promuovere i diritti umani di fronte all'utilizzo di sistemi che sfruttano le tecniche di Intelligenza Artificiale.**



Campo di applicazione



La Raccomandazione dell'UNESCO affronta l'etica dell'Intelligenza Artificiale come una riflessione multiculturale e in evoluzione di valori, principi e azioni interdipendenti (in ottica di normativa sistematica olistica) che possano aiutare e guidare le società nell'affrontare in modo responsabile gli impatti noti e sconosciuti delle tecnologie legate all'alveo dell'Intelligenza Artificiale sugli esseri umani, sulle società, sull'ambiente e sugli ecosistemi, offrendo così una solida base per comprendere, accettare o rifiutare le tecnologie dell'IA.



In questo documento si enfatizza la natura dinamica dell'etica stessa, che cambia geograficamente e cambia nel tempo. Tuttavia, ci sono alcuni principi etici comuni, a livello globale, che possono - e devono - fare da bussola per tutti i Paesi. In quanto strumento di definizione degli standard sviluppato attraverso un approccio globale basato sul diritto internazionale, la Raccomandazione UNESCO si concentra, infatti, su principi etici condivisi a livello globale quali la dignità umana e i diritti umani, così come l'uguaglianza di genere, la giustizia sociale ed economica e lo sviluppo, il benessere fisico e mentale, la diversità, l'interconnessione, l'inclusività, e la protezione dell'ambiente e dell'ecosistema.



- Partendo da questi valori condivisi, l'approccio intrapreso dall'UNESCO mira a stimolare l'attenzione (nonché l'azione concreta) su alcune specifiche caratteristiche dei sistemi di Intelligenza Artificiale che sono di centrale rilevanza etica, ponendo l'accento su tre distinte direttrici:
 - 1 — i sistemi di IA legati all'elaborazione delle informazioni che integrano modelli e algoritmi che producono una capacità di apprendere ed eseguire compiti cognitivi - con differenti livelli di autonomia delle macchine - che portano a risultati come la previsione/pianificazione e il processo decisionale, tanto in ambienti materiali/fisici quanto in ambienti virtuali;
 - 2 — le questioni etiche devono sempre riguardare tutte le fasi del ciclo di vita di un sistema basato su Intelligenza Artificiale. Ciclo di vita che va dalla ricerca, alla progettazione e sviluppo, fino al rilascio e distribuzione, all'uso, compresa la manutenzione, il funzionamento, il commercio, il finanziamento, il monitoraggio e la valutazione, la convalida, la fine dell'utilizzo, la cessazione;
 - 3 — i sistemi di IA sollevano nuovi tipi di questioni etiche, che includono il loro impatto sul processo decisionale, l'occupazione e il lavoro, l'interazione sociale, l'assistenza sanitaria, l'istruzione, i media, l'accesso alle informazioni, il divario digitale, i dati personali e la protezione dei consumatori, l'ambiente, la democrazia, lo stato di diritto, la sicurezza e la polizia, i diritti umani e le libertà fondamentali, compresa la libertà di espressione, la privacy e la non discriminazione. Inoltre, nuove sfide etiche sono create dal potenziale degli algoritmi di IA di riprodurre e rafforzare i pregiudizi esistenti e, quindi, di esacerbare forme già esistenti di discriminazione, pregiudizio e stereotipi.



Ci soffermiamo su quest'ultimo punto, perché troviamo particolarmente interessanti le riflessioni riportate nel testo della Raccomandazione sull'etica dell'Intelligenza Artificiale dell'UNESCO, che qui citiamo:

«Alcuni di questi problemi sono legati alla capacità dei sistemi di IA di eseguire compiti che in precedenza solo gli esseri viventi potevano fare e che, in alcuni casi, erano addirittura limitati ai soli esseri umani. Queste caratteristiche danno ai sistemi di IA un ruolo nuovo e profondo nelle pratiche umane e nella società, così come nel rapporto con l'ambiente e gli ecosistemi, creando un nuovo contesto in cui i bambini e i giovani possono crescere, sviluppare una comprensione del mondo e di sé stessi, comprendere criticamente i media e le informazioni e imparare a prendere decisioni. A lungo termine, i sistemi di IA potrebbero sfidare il senso speciale dell'esperienza e dell'azione degli esseri umani, sollevando ulteriori preoccupazioni riguardo, tra l'altro, all'autocomprensione umana, all'interazione sociale, culturale e ambientale, all'autonomia, all'azione, al valore e alla dignità».





Scopi e obiettivi

La Raccomandazione dell'UNESCO mira a portare uno strumento normativo globalmente accettato che si concentra non solo sull'articolazione dei valori e dei principi, ma anche sulla loro realizzazione pratica, attraverso raccomandazioni politiche concrete.

Poiché la complessità delle questioni etiche che circondano l'IA richiede la cooperazione di più parti interessate nelle comunità internazionali, regionali e nazionali, la Raccomandazione punta a consentire a tutte queste parti di assumere una responsabilità condivisa basata su un dialogo globale e interculturale. **Questi gli obiettivi che si sono dati i 193 Paesi che hanno sottoscritto la carta:**

- 1 — **fornire un quadro universale di valori**, principi e azioni per guidare gli Stati nella formulazione delle loro legislazioni, politiche o altri strumenti riguardanti l'IA, coerentemente con il diritto internazionale;
- 2 — **guidare le azioni** di individui, gruppi, comunità, istituzioni e aziende del settore privato per assicurare l'integrazione dell'etica in tutte le fasi del ciclo di vita dei sistemi basati su Intelligenza Artificiale;
- 3 — **proteggere, promuovere e rispettare i diritti umani e le libertà fondamentali**, la dignità umana e l'uguaglianza, compresa la parità di genere; salvaguardare gli interessi delle generazioni presenti e future; preservare l'ambiente, la biodiversità e gli ecosistemi; rispettare la diversità culturale in tutte le fasi del ciclo di vita delle tecnologie;
- 4 — **promuovere il dialogo multi-stakeholder**, multidisciplinare e pluralistico e la costruzione del consenso sulle questioni etiche relative ai sistemi di IA;
- 5 — **promuovere un accesso equo agli sviluppi e alle conoscenze nel campo dell'IA** e la condivisione dei benefici, con particolare attenzione ai bisogni e ai contributi dei paesi meno sviluppati.





Valori e principi


La parte più corposa della Raccomandazione riguarda i capitoli dedicati a valori e principi, in quanto - si legge nel documento - «*giocano un ruolo potente come ideali motivanti nel plasmare misure politiche e norme legali*».




I valori condivisi a livello globale sui quali si dovranno fondare le singole scelte degli Stati membri, possono essere riassunti così:

-  **rispetto, protezione e promozione:** dei diritti umani, delle libertà fondamentali e della dignità umana

-  **attenzione:** all'ambiente e agli ecosistemi, non solo riducendo l'impatto ambientale ma anche stimolando ecosistemi (anche economici e sociali) fiorenti


-  **garanzia:** non solo del rispetto ma anche della promozione di diversità e inclusione


-  **ruolo partecipativo e abilitante:** per garantire società pacifiche e giuste, basate su un futuro interconnesso a beneficio di tutti





I principi da tenere bene in considerazione affinché tali valori vengano rispettati lungo tutto il ciclo di vita dei sistemi basati su Intelligenza Artificiale, sono i seguenti:


+

 **proporzionalità e non nocività:** il metodo e le tecniche di IA scelte dovrebbero essere appropriati e proporzionali per raggiungere un determinato scopo legittimo, non dovrebbero mai violare i valori fondamentali riportati nel testo di Raccomandazione ed essere appropriati al contesto, nonché basati su basi scientifiche rigorose;

 **sicurezza e protezione:** l'IA sicura sarà abilitata dallo sviluppo di strutture di accesso ai dati sostenibili e protettive della privacy che favoriscano una migliore formazione e validazione dei modelli di IA utilizzando dati di qualità;


 **equità e non discriminazione:** tutti i player attivi sul fronte IA (dal mondo della ricerca alle aziende tecnologiche private) dovranno fare tutti gli sforzi ragionevoli per minimizzare ed evitare di rinforzare o perpetuare applicazioni e risultati discriminatori o distorti durante tutto il ciclo di vita di un sistema di IA per assicurare l'equità di tali sistemi. Si auspica un rimedio efficace contro la discriminazione e la determinazione algoritmica distorta;


 **sostenibilit:** la valutazione continua dell'impatto umano, sociale, culturale, economico e ambientale delle tecnologie di IA dovrebbe essere effettuata con piena conoscenza delle implicazioni delle tecnologie di IA per la sostenibilità come un insieme di obiettivi in costante evoluzione attraverso una serie di dimensioni, come attualmente identificato negli Obiettivi di sviluppo sostenibile (SDGs) delle Nazioni Unite;


 **diritto alla privacy e alla protezione dei dati:** la privacy viene riconosciuta come un diritto essenziale per la protezione della dignità e dell'autonomia umana; dovrà quindi essere rispettata, protetta e promossa durante tutto il ciclo di vita dei sistemi di IA. Allo stesso tempo, adeguati quadri normativi di protezione





dei dati e meccanismi di governance dovrebbero essere stabiliti in un approccio multi-stakeholder a livello nazionale e internazionale, nonché protetti e garantiti da sistemi giudiziari;

 **supervisione umana per garantire responsabilità etica e legale:** gli Stati membri dovrebbero garantire che sia sempre possibile attribuire la responsabilità etica e legale per qualsiasi fase del ciclo di vita dei sistemi di IA;

 **trasparenza e spiegabilità:** trasparenza e spiegabilità dei sistemi di IA sono precondizioni essenziali per garantire il rispetto, la protezione e la promozione dei diritti umani, delle libertà fondamentali e dei principi etici. La trasparenza è necessaria affinché i regimi di responsabilità nazionali e internazionali pertinenti funzionino efficacemente. Una mancanza di trasparenza potrebbe anche minare la possibilità di contestare efficacemente le decisioni basate sui risultati prodotti dai sistemi di IA e potrebbe quindi violare il diritto a un processo equo e a un ricorso efficace. Sarà dunque necessario aumentare al massimo gli sforzi in questa direzione che, attualmente, è una delle aree più critiche da gestire;

 **responsabilità e rendiconto:** la responsabilità etica e la responsabilità per le decisioni e le azioni basate in qualsiasi modo su un sistema di IA dovrebbero essere sempre, in ultima analisi, attribuibili agli attori coinvolti nel ciclo di vita del sistema di IA. A tal fine, dovranno essere sviluppati meccanismi appropriati di supervisione, valutazione degli impatti, audit e due diligence;

 **consapevolezza e alfabetizzazione:** la consapevolezza e la comprensione pubblica delle tecnologie e il valore dei dati dovrebbero essere promossi attraverso un'educazione aperta e accessibile, l'impegno civico, le competenze digitali e la formazione sull'etica dell'IA;

 **governance e collaborazione adattiva:** nella Raccomandazione si enfatizza il ruolo importante della sovranità nazionale per il rispettato nell'uso dei dati, ma si parla anche di apertura e collaborazione adattiva. La partecipazione di diversi stakeholder in tutto il ciclo di vita dei sistemi di IA è necessaria per approcci inclusivi alla governance dell'IA, permettendo che i benefici siano condivisi da tutti e che contribuiscano allo sviluppo sostenibile.





Governance e azioni politiche

I capitoli finali della Raccomandazione dell'UNESCO sono dedicati alla Governance e ai suggerimenti delle azioni politiche che ciascuno Stato membro dovrebbe seguire per dare concretezza a quanto riportato nel documento stesso. Gli Stati membri dovrebbero introdurre quadri regolatori per le valutazioni d'impatto, come la valutazione dell'impatto etico, per identificare e valutare i benefici, le preoccupazioni e i rischi dei sistemi di IA, così come le appropriate misure di prevenzione, mitigazione e monitoraggio dei rischi, tra gli altri meccanismi di garanzia.



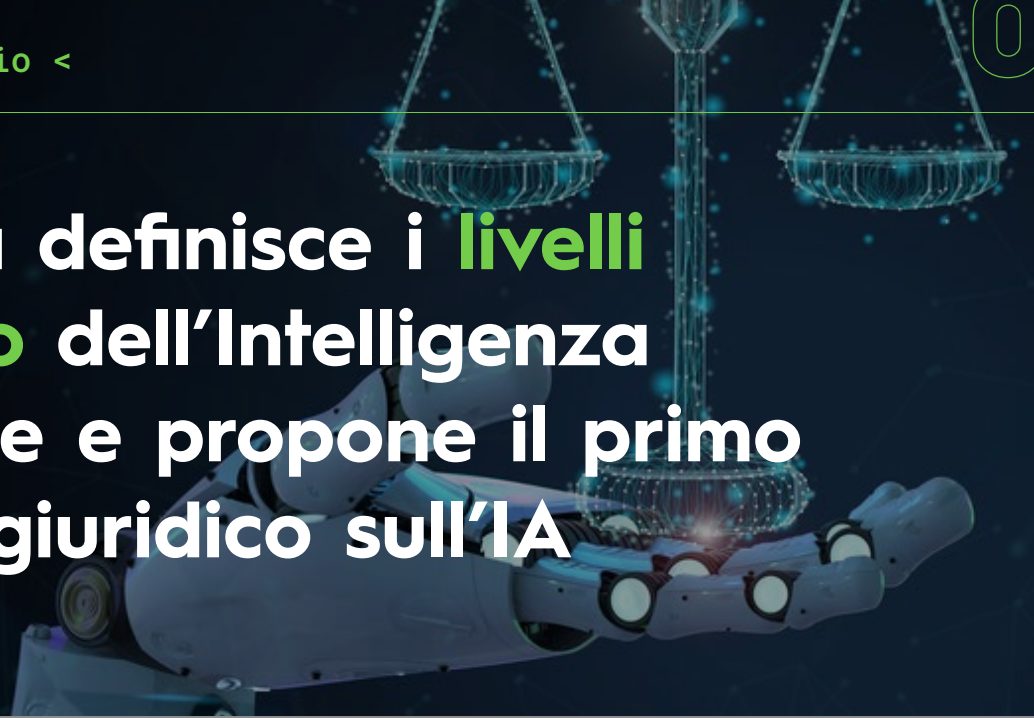
In quest'ottica, anche le aziende del settore privato dovrebbero sviluppare meccanismi di due diligence e di supervisione per identificare, prevenire, mitigare e rendere conto di come affrontano l'impatto dei sistemi di IA sul rispetto dei diritti umani, sullo stato di diritto e sulle società inclusive.



La Raccomandazione dell'UNESCO invita, di fatto, gli Stati membri a coinvolgere tutte le parti interessate, comprese le imprese commerciali, per garantire che svolgano i loro rispettivi ruoli del raggiungimento degli obiettivi definiti nel documento portandolo all'attenzione delle autorità, degli enti, delle organizzazioni di ricerca e accademiche, delle istituzioni e delle organizzazioni nei settori pubblico, privato e della società civile coinvolti nelle tecnologie di IA, in modo che lo sviluppo e l'uso delle tecnologie basate su Intelligenza Artificiale siano guidati sia da una solida ricerca scientifica sia da analisi e valutazioni etiche, riconoscendo che affrontare i rischi e le preoccupazioni etiche non rappresenta affatto un ostacolo all'innovazione e allo sviluppo, ma fornisce piuttosto nuove opportunità perché siano condotte in modo etico.



L'Europa definisce i **livelli di rischio** dell'Intelligenza Artificiale e propone il primo quadro giuridico sull'IA



Dopo la "Dichiarazione di cooperazione sull'Intelligenza Artificiale" dell'aprile 2018; le "Linee guida etiche finali per un'Intelligenza Artificiale affidabile" dell'aprile 2019; il "Rapporto sulla responsabilità per l'Intelligenza Artificiale e altre tecnologie emergenti" del novembre 2019 e il "Libro bianco sull'Intelligenza Artificiale" del febbraio 2020, il 21 aprile 2021 la Commissione europea ha pubblicato la proposta di un Regolamento volto a definire uno standard comune per lo sviluppo di un'IA sicura, di cui gli utenti possano fidarsi, che poggi su una tecnologia etica, nel pieno rispetto dei valori e dei diritti fondamentali dell'Unione.

Definendo precisi livelli di rischio per differenti sistemi e applicazioni di Intelligenza Artificiale, predisponendo misure di sorveglianza per i sistemi "ad alto rischio" e vietando alcune pratiche di IA perché contrarie ai principi dell'UE, il nuovo Regolamento, di fatto, getta le basi del primo quadro giuridico europeo sull'Intelligenza Artificiale.

A questo si aggiungono (sempre datati aprile 2021) il nuovo Piano coordinato sull'Intelligenza Artificiale 2021 - che delinea i cambiamenti politici e gli investimenti atti a rendere più forte la posizione dell'Europa nello sviluppo di un'IA incentrata sull'uomo, sostenibile, sicura, e affidabile - e il nuovo Regolamento Macchine, che adatta le norme di sicurezza al bisogno di fiducia, da parte degli utenti, nei confronti di una nuova generazione di tecnologie e di prodotti.







Quattro livelli di rischio

*



Il nuovo Regolamento sarà applicato allo stesso modo in tutti gli Stati membri, sulla base di una definizione di Intelligenza Artificiale che segue un approccio basato sul “rischio”. Più nel dettaglio, la Commissione europea ha classificato quattro livelli di rischio, in riferimento a specifici sistemi di Intelligenza Artificiale e ai loro ambiti di applicazione:







-  **il rischio inaccettabile** rimanda a sistemi IA considerati una «chiara minaccia alla sicurezza degli utenti», tra cui quelle applicazioni tendenti a manipolare il comportamento delle persone, annullando il loro libero arbitrio, o a sviluppare un sistema nazionale per classificare la reputazione dei cittadini sulla base di informazioni riguardanti la loro condizione economica e sociale;
-  **il rischio elevato** fa riferimento a sistemi e tecnologie IA utilizzati nell’ambito di:
 - infrastrutture critiche
 - formazione scolastica e professionale
 - componenti di sicurezza dei prodotti
 - occupazione e gestione dei lavoratori
 - servizi privati e pubblici (in cui, ad esempio, un algoritmo IA potrebbe sbarrare la strada all’ottenimento di un prestito)
 - Forze dell’Ordine
 - gestione della migrazione, dell’asilo politico e dei controlli alle frontiere
 - amministrazione della giustizia e dei processi
-  **il rischio limitato** riguarda, invece, quei sistemi di Intelligenza Artificiale - come, ad esempio, i chatbot - in merito ai quali gli utenti devono essere consapevoli del fatto che stanno interagendo con una macchina, in modo da scegliere in totale autonomia se proseguire oppure fare un passo indietro;
-  **il rischio minimo** è - solo per citare alcuni esempi - quello dei videogiochi abilitati all’Intelligenza Artificiale o dei filtri anti-spam. La grande maggioranza dei sistemi di Intelligenza Artificiale rientra in questa categoria. E il progetto per l’Intelligenza Artificiale in Europa non interviene in questi casi, poiché tali sistemi rappresentano un rischio giudicato “minimo” o “nullo” per i diritti e la sicurezza dei cittadini.





Misure di sorveglianza, obblighi e divieti

La novità introdotta dalla proposta di Regolamento europeo per un'IA sicura sta negli obblighi ai quali i sistemi di Intelligenza Artificiale classificati "a rischio elevato" saranno soggetti prima di poter essere immessi sul mercato. Obblighi che prevedono:

-  **sistemi di valutazione del rischio;**
.....
-  **elevata qualità dei set di dati che andranno ad addestrare il sistema IA**, col fine di minimizzare i rischi di discriminazione;
.....
-  **registrazione delle attività**, per garantire la tracciabilità dei risultati;
.....
-  **documentazione dettagliata**, che fornisce tutte le informazioni necessarie sul sistema e sul suo scopo, affinché le Autorità possano valutarne la conformità;
.....
-  **informazioni chiare da fornire all'utente;**
.....
-  **adeguate misure di sorveglianza da parte di figure preposte.**

In particolare, il nuovo Regolamento considera a rischio elevato - e soggetti a severi requisiti - tutti i sistemi di identificazione biometrica remota, il cui utilizzo dal vivo, all'interno di spazi pubblici, è in linea di principio vietato.

Le eccezioni devono essere rigorosamente definite e regolate. Queste comprendono le ricerche di soggetti minori scomparsi, le prevenzioni di eventuali minacce terroristiche e l'individuazione degli autori reati ritenuti gravi. E tale utilizzo - precisa il documento della Commissione europea - è comunque

«soggetto all'autorizzazione di un organo giudiziario o di altro organo indipendente e a opportuni limiti di tempo, di portata geografica e di banche dati utilizzate per la ricerca».



La proposta di Regolamento, infine, introduce un elemento importante relativo alla governance dei sistemi di Intelligenza Artificiale ad alto rischio, proponendo che, in tutti gli Stati membri, le Autorità nazionali di vigilanza prepose si impegnino affinché misure, obblighi e divieti vengano rispettati e che venga istituito un apposito Comitato europeo per l'Intelligenza Artificiale, allo scopo di "facilitare" l'attuazione delle nuove regole.





Il nuovo Regolamento Macchine



Come accennato, alla proposta di nuove regole per disciplinare l'Intelligenza Artificiale in Europa, si è aggiunto il nuovo Regolamento Macchine, che va a sostituire la ben nota Direttiva Macchine in tema di requisiti di sicurezza.

I prodotti delle macchine - lo ricordiamo - coprono una vasta gamma di dispositivi di consumo e di apparecchiature professionali, dai robot alle stampanti 3D, dalle macchine edili alle linee di produzione industriale, solo per citare alcuni esempi.

Mentre la proposta di Regolamento sull'Intelligenza Artificiale affronta la spinosa tematica concernente i rischi e la sicurezza dei sistemi IA,

il nuovo Regolamento Macchine garantirà l'integrazione "sicura" dei sistemi di Intelligenza Artificiale nel macchinario complessivo.



A tale riguardo, le aziende, rispetto al passato, dovranno eseguire una sola valutazione di conformità.



Inoltre, il Regolamento risponderà alle nuove esigenze del mercato, apportando maggiore chiarezza giuridica alle disposizioni attuali, semplificando l'onere amministrativo e i costi per le aziende, consentendo formati digitali per la documentazione e adattando le tariffe di valutazione della conformità per le PMI, garantendo - al contempo - la coerenza con l'UE per quanto concerne il quadro legislativo per i prodotti.



La **strategia italiana** per un'IA antropocentrica, affidabile e sostenibile

L'UNESCO rilascia il documento di Raccomandazione sull'etica dell'Intelligenza Artificiale il 25 novembre 2021 ed esattamente un giorno prima, il 24 novembre 2021, il nostro Consiglio dei Ministri approva il Programma Strategico per l'Intelligenza Artificiale 2022-2024, a cura del Ministero dell'Università e della Ricerca, del Ministero dello Sviluppo Economico (MiSE) e del Ministro per l'innovazione tecnologica e la transizione digitale, contenente, tra gli obiettivi, anche quello relativo allo sviluppo di un'Intelligenza Artificiale antropocentrica, affidabile e rispettosa delle normative, tra cui quella in tema di protezione dei dati personali.

Ricordiamo che, in Italia, si fa riferimento per la prima volta - in seno a un documento istituzionale - alla **“visione antropocentrica”** dell’IA nelle proposte per una “Strategia italiana per l’Intelligenza Artificiale”, elaborate a luglio del 2020 dal gruppo di esperti del MiSE in risposta al “Piano Coordinato” della Commissione Europea in materia di Intelligenza Artificiale.

«Un approccio antropocentrico all’Intelligenza Artificiale richiede che l’essere umano venga valorizzato e supportato dalla tecnologia nel suo realizzarsi come individuo, nel suo agire come cittadino, nelle sue mansioni e competenze come lavoratore e nelle sue decisioni come consumatore»

si legge nel documento del 2020. E, ancora:






«L’Intelligenza Artificiale deve essere adottata come metodo educativo in grado di portare alla formazione del cosiddetto pensiero computazionale, alla multidisciplinarietà intrinseca nella soluzione dei problemi e nella trasversalità delle competenze».

Si tratta di una sfida niente affatto semplice, che richiede l’apporto di molteplici ambiti di intervento e di iniziative, dall’aggiornamento degli educatori e del corpo docente all’integrazione dei percorsi di studi con corsi di coding e di Intelligenza Artificiale applicata, fino all’up-skilling e re-skilling della forza lavoro, passando per informazione, divulgazione, formazione dell’attuale management aziendale e della classe imprenditoriale.



I principi guida del Programma per l'Intelligenza Artificiale 2022-2024

Quella espressa dal Programma italiano per il prossimo triennio è una strategia per l'Intelligenza Artificiale che ricalca il piano dell'UE in materia, ispirandosi, nel dettaglio, ai seguenti principi guida:

-  **seguire la linea europea**, focalizzandosi, per quanto concerne l'Intelligenza Artificiale, «sulla cooperazione, sui dati e sull'infrastruttura informatica» e identificando i settori prioritari, all'interno dei quali costruire una propria leadership;
-  **puntare a divenire polo di ricerca e innovazione**, rafforzando gli investimenti in ricerca e sviluppo, con l'obiettivo di mettere a punto tecniche innovative in tema di Intelligenza Artificiale;
-  **adottare l'approccio per un'IA antropocentrica, affidabile e sostenibile**, ossia basare lo sviluppo dell'Intelligenza Artificiale sui diritti e sui valori umani, nonché sui concetti di sostenibilità ambientale, sociale ed economica. In Italia, la sicurezza dell'IA, la sua progettazione responsabile e trasparente, devono essere in cima alle priorità, in tutti i settori. A tale riguardo, ricordiamo che il riferimento è dato dalle "Linee guida etiche per un programma di orientamento e attuazione affidabile dell'IA", definito dall'High Level Expert Group on AI istituito dalla Commissione Europea;
-  **intervenire sulla trasformazione digitale del tessuto imprenditoriale**, accelerando l'adozione di soluzioni IA in tutte le aziende italiane;
-  **governare l'IA e governare "con" l'IA**, in riferimento - da un lato - all'impegno, da parte delle Pubbliche Amministrazioni del nostro Paese, nel mitigare tutti i possibili rischi correlati all'utilizzo dell'Intelligenza Artificiale e nel porre sempre in prima linea i diritti umani e il rispetto dei principi etici e - dall'altro - nel fare dell'IA uno strumento "responsabile", atto a modernizzare operatività, processi e politiche.





Gli obiettivi da conseguire entro il prossimo triennio

In particolare, uno dei punti sul quale insiste il Programma Strategico per l'Intelligenza Artificiale 2022-2024 è la **crucialità dell'attività di ricerca e sviluppo**, base sulla quale poggiano un'IA innovativa, lo sviluppo del mercato e la crescita economica. Dunque, gli investimenti nella ricerca sono al primo posto della nuova strategia italiana per l'artificial intelligence.



E sempre in linea con l'approccio europeo - che vuole un'Intelligenza Artificiale «antropocentrica, affidabile e sostenibile» - nonché con gli obiettivi per lo sviluppo sostenibile delle Nazioni Unite, il nuovo Programma nazionale sostiene un'IA che non solo supporta la **crescita economica**, ma fa in modo che questa sia «inclusiva e sostenibile».

Un altro aspetto che viene rimarcato in più passaggi del Programma riguarda, inoltre, l'imperativo della **trasformazione digitale di tutto l'ecosistema imprenditoriale italiano** e, parallelamente, delle Pubbliche Amministrazioni, con la spinta all'implementazione e all'adozione - in questi settori - di soluzioni di Intelligenza Artificiale.



Ricapitolando, gli obiettivi da conseguire entro il 2024 sono, in totale, sei e prevedono - sempre a proposito della ricerca sull'IA - l'adozione di un approccio multidisciplinare e la riduzione della tendenza alla frammentazione, favorendo, la collaborazione tra segmenti diversi.

Il terzo obiettivo fa riferimento - come accennato - allo sviluppo di un'Intelligenza Artificiale che ponga al centro l'essere umano, antropocentrica e affidabile, rispettosa delle normative.

E «aumentare l'innovazione basata sull'IA e lo sviluppo della tecnologia di IA», sia nelle PMI che nel settore pubblico, rappresentano altri due obiettivi previsti dalla nuova strategia italiana.

Infine, l'ultimo obiettivo:

«creare una nuova generazione di ricercatori e di innovatori in ambito IA, col fine di rendere l'Italia una meta attrattiva per ricercatori qualificati esteri, mantenendo una particolare attenzione alla diversità e all'equilibrio di genere».



I settori sui quali investire e le politiche previste

Industria e manifatturiero; sistema educativo; agroalimentare; cultura e turismo; salute e benessere; ambiente, infrastrutture e reti; banche, finanza e assicurazioni; Pubblica Amministrazione; città, aree e comunità intelligenti; sicurezza nazionale; tecnologie dell'informazione: sono questi i settori sui quali il nostro Paese è pronto a investire in IA, da qui al 2024.

Interessante è l'attenzione rivolta al **sistema educativo** che, dal punto di vista tecnologico, è stato, negli ultimi anni, spesso fanalino di coda. La nuova strategia italiana per l'Intelligenza Artificiale, al riguardo, prevede, invece, un nuovo piano di istruzione per "formare" alla tecnologia - e, più in particolare, all'IA - e la messa a punto, grazie ai sistemi di artificial intelligence, di programmi di apprendimento personalizzati.

Quello della **salute e benessere** è un settore che, a partire dagli ultimi cinque anni - e soprattutto con l'emergenza pandemica - ha visto un grande sforzo da parte della ricerca e un numero elevato di applicazioni di IA. Il nuovo piano strategico intende proseguire su questa strada, privilegiando in modo particolare l'adozione di tecniche di Intelligenza Artificiale per lo studio e la diagnosi precoce delle malattie degenerative, delle patologie tumorali e virali, lo sviluppo di nuovi farmaci e vaccini, il monitoraggio dei pazienti e la predittività delle emergenze sanitarie.

Per quanto concerne invece, il **settore banche, finanza e assicurazioni**, gli investimenti in IA che il Programma strategico contempla sono finalizzati a elevare la qualità e il livello di sicurezza dei servizi ai clienti e, al contempo, a ridurne i

costi. Un altro obiettivo è quello di potenziare i meccanismi di prevenzione delle frodi per mezzo di sistemi di Intelligenza Artificiale sviluppati ad hoc.

Tra i settori prioritari sui quali investire, anche quello afferente a **città, aree e comunità intelligenti**, alle quali le tecniche di Intelligenza Artificiale - si legge nel Programma - permetteranno di accedere con facilità a tutti i servizi messi a disposizione dei cittadini, ovunque vivano, città o area rurale. Un altro ambito di intervento è quello che mira a ridurre il traffico nelle aree cittadine (e il livello di inquinamento che ne deriva) grazie a soluzioni IA quali il parcheggio intelligente, i sistemi di gestione del traffico e i veicoli a guida autonoma.

Quella dei **talenti e delle competenze** rappresenta la prima area di intervento suggerita dal Programma italiano, in cui trovano posto una serie di politiche focalizzate sullo sviluppo di capitale umano con saperi e abilità nelle IA. Serve personale preparato sulla materia, con competenze in tutti i suoi segmenti, ma anche cittadini con maggiori conoscenze in tema di IA, si legge nel Documento. Tra le azioni, in tale area, per i prossimi tre anni, l'ampliamento dei programmi di dottorato e il potenziamento della presenza delle materie STEM in tutti i moduli di istruzione del nostro Paese.

L'area della ricerca - come già, in parte, illustrato - prevede azioni e iniziative volte a rafforzare l'ecosistema italiano della ricerca sull'IA, a creare cattedre di ricerca sulla materia, nonché a stimolare sinergie e partnership tra Università, mondo dell'industria, PMI, Enti pubblici e semplici cittadini.

L'area delle applicazioni, infine, consta di politiche che puntano all'ingresso delle tecniche di Intelligenza Artificiale negli undici settori prioritari individuati. L'obiettivo - come già sottolineato - è uno solo: modernizzare, spingere verso la trasformazione digitale imprese e PA italiane, con l'occhio attento soprattutto a quelle realtà più piccole e provenienti da aree periferiche del Paese e a quei contesti in cui è alto il rischio di una minore presenza femminile.

Perché nessuno deve rimanere indietro, né sentirsi escluso: questo è il fine ultimo del nuovo Programma italiano per l'IA.

Lo “Statuto etico e giuridico dell’Intelligenza Artificiale” di Fondazione Leonardo-Civiltà delle Macchine

L’Intelligenza Artificiale, il suo essere presente ormai nel nostro quotidiano, il suo interagire con la vita economica e sociale del Paese, pone interrogativi importanti, in cui la domanda non è tanto “come funziona l’IA?”, ma “come si comporta l’IA?”. È il pensiero di Fondazione Leonardo-Civiltà delle Macchine, per il quale un “Umanesimo digitale” si impone, guidato dalla «necessità che i sistemi di Intelligenza Artificiale si conformino a un approccio antropocentrico, a beneficio della collettività. Le innovazioni devono essere programmate mettendo al centro l’uomo, il rispetto dei suoi diritti e dei valori propri di una società democratica».

Seguendo tale indirizzo, la Fondazione ha messo a punto lo “Statuto etico e giuridico dell’Intelligenza Artificiale” che, accorpando tre documenti (a cura di esperti in materia) su questioni di carattere tecnico, etico e giuridico, **intende offrire uno schema di good practice all’industria dei settori IA e robotica**, inquadrando i nodi etici della società digitale e proponendo ai decisori politici un insieme di regole per garantire che le nuove tecnologie restino una forza benefica all’interno della nostra società.



La centralità dell'educazione alle nuove tecnologie: l'IA è di tutti e per tutti

«L'IA non è un terreno destinato a caste ristrette. È la madre delle grandi innovazioni del nostro tempo, coinvolge l'intera umanità e sfida il presente in termini radicali»

+



J

si legge nello Statuto Etico e Giuridico dell'Intelligenza Artificiale. Si tratta di una posizione di rottura rispetto al passato, quando i meccanismi e le dinamiche che fungono da base all'Intelligenza Artificiale venivano tenute sotto una teca di vetro, con le chiavi custodite dagli addetti ai lavori. Sapere esclusivo di ingegneri, informatici e sviluppatori, l'IA e, più in generale, le tecnologie digitali, restavano "materia oscura" per i comuni cittadini.

«Dalle caste ristrette all'intera umanità» recita, invece, lo Statuto della Fondazione. Perché gli utenti - in ambito lavorativo come nel privato - sono i cittadini. Ai quali, oggi, viene chiesto non solo di conoscere l'IA, la robotica e le nuove tecnologie, ma di essere consapevoli di come queste possano agire, di come possano comportarsi, nel bene e nel male.



In tale dibattito, la questione della formazione diviene centrale e cruciale: è necessaria un'alfabetizzazione digitale che insegni al cittadino comune l'uso dell'Intelligenza Artificiale e della robotica, una formazione degli utenti che permetta un



utilizzo consapevole delle nuove tecnologie, che ne evidenzi le potenzialità, responsabilizzando l'individuo di fronte ai rischi.

Rischi che possono essere connessi - solo per citare alcuni esempi - a un uso doloso dell'IA, intenzionalmente guidato da scopi illegali (attacchi informatici, manipolazione dell'opinione pubblica attraverso fake news, campagne di disinformazione), oppure a un suo utilizzo non doloso, tra cui il pregiudizio non intenzionale e la discriminazione; rischi che hanno a che vedere con il corretto funzionamento degli algoritmi o con l'interazione tra uomo e macchina (che andrà, col tempo, a influire su determinate dinamiche comportamentali), rischi legati ai requisiti di sicurezza dei materiali con i quali vengono costruiti i robot e rischi causati da comportamenti non regolamentati da parte di questi.





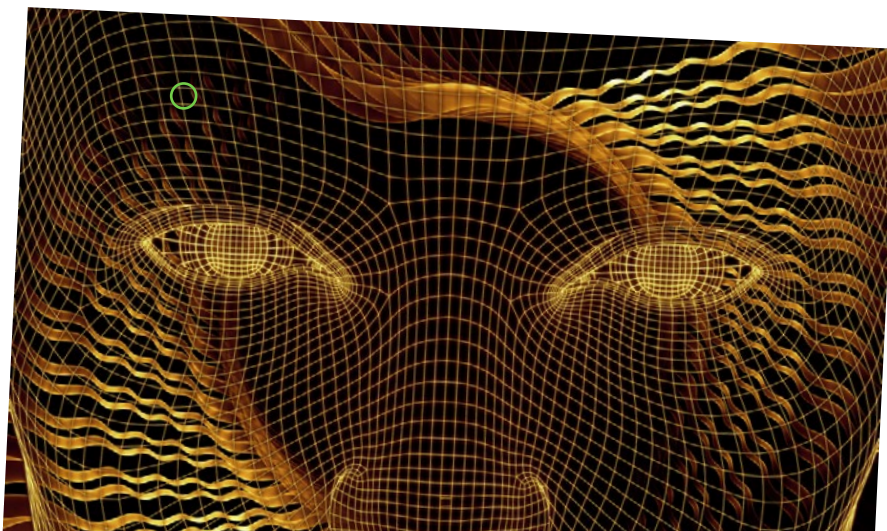
I principi e i valori per un uso etico dell'Intelligenza Artificiale



In particolare, per quanto concerne il documento relativo alle questioni di carattere etico, lo Statuto di Fondazione Leonardo-Civiltà delle Macchine identifica alcuni principi e valori (declinati in tre macro livelli: individuale, sociale e globale) che le tecniche di Intelligenza Artificiale non dovranno in alcun modo ledere. Appartengono al primo livello i valori propri della persona, di cui la **dignità umana** rappresenta quello nodale, da cui derivano **libertà, diritti civili e il principio della non discriminazione**.

Agli antipodi rispetto alla dignità umana, sono - ad esempio - quelle tecnologie che tendono a “manipolare” la persona o alle quali vengono affidate scelte e decisioni rilevanti per l'esistenza del soggetto, ma di cui non è dato conoscerne i meccanismi e i passaggi.

Fanno riferimento al livello sociale, i principi **di inclusività, di riduzione della disuguaglianza e di coesione sociale**, «da rispettare e promuovere, in un'ottica di bilanciamento tra bene comune e individuale». Qui l'accento è posto sul concetto di “equità”, sposando il quale, l'adozione di sistemi IA deve poter assicurare a tutti pari opportunità derivanti dall'accesso ai servizi che ne derivano.



Secondo Fondazione Leonardo-Civiltà delle Macchine,

«l'IA può diventare una forza attiva per la riduzione delle diseguaglianze, incorporando un concetto di giustizia distributiva che guardi alle categorie marginali come soggetti di intervento prioritario: strumenti basati sull'IA, ad esempio, possono risultare utili, nell'ambito del sistema educativo, per colmare divari di apprendimento mentre, in ambito sanitario, possono essere utilizzati per stimolare il social empowerment e i servizi destinati agli individui affetti da disabilità».

Infine, a livello globale, sono la **prevenzione del danno, pace e giustizia e sostenibilità** i valori per un utilizzo etico dell'Intelligenza Artificiale. Nel dettaglio, il primo di tali valori presuppone un'attenta analisi dei rischi, volta a prendere in considerazione determinate soluzioni che proteggano le persone dal loro manifestarsi. Quella della prevenzione è una prassi ormai consolidata all'interno della maggior parte delle aziende (si pensi solo alla sicurezza dei beni materiali e alla sicurezza informatica) e che - si legge nello Statuto - andrebbe estesa anche ai sistemi IA, valutandone, ad esempio, l'eventuale trasmissione di bias ai dati utilizzati per l'addestramento degli algoritmi.

Infine, quando si parla di sviluppo etico di sistemi di Intelligenza Artificiale si deve fare necessariamente riferimento - secondo Fondazione Leonardo-Civiltà delle Macchine - a macchine che non mettano a rischio i valori di pace e giustizia,

«di concordia e fratellanza, contribuendo a prevenire, anche grazie all'innovazione tecnologica, lo scoppio di conflitti e il concretizzarsi di tensioni internazionali».



Il rispetto dei diritti fondamentali della persona

Informazione, educazione, autodeterminazione dell'identità, riservatezza, tutela - in generale - dei propri diritti e, in particolare, dei diritti dei soggetti più deboli: sono i diritti della persona che lo Statuto pone in evidenza e che devono essere rispettati nell'ambito dello sviluppo e dell'utilizzo dei sistemi di Intelligenza Artificiale. Vediamoli, in breve sintesi, uno ad uno:

- ☀ **diritto ad essere informato** da parte di chi utilizza un sistema IA che preveda uno scambio di dati - circa il processo che va dalla raccolta delle informazioni alla loro elaborazione, fino all'analisi di eventuali rischi correlati;
- ☀ **diritto ad essere educato e formato** sul funzionamento e sul corretto utilizzo di un determinato sistema IA, nonché sui suoi potenziali rischi, in modo da esserne consapevole;
- ☀ **diritto all'autodeterminazione dell'identità** nei casi in cui il soggetto entri in contatto con un sistema di Intelligenza Artificiale tendente a «manipolare le sue decisioni, minando, così, il diritto a scegliere e a decidere per se stesso». L'informazione è, in questo caso, basilare per capire come interagire con la macchina preservando la propria libertà;
- ☀ **diritto alla riservatezza dei propri dati personali**, ceduti alla macchina con finalità predittive, decisionali o relative alla deduzione di nuove informazioni, col rischio di impatti negativi sulla sfera privata della persona;
- ☀ **tutela, in generale, dei diritti della persona**, con l'attenzione, in particolare, alle specifiche caratteristiche e alle funzioni svolte dal sistema IA in questione;
- ☀ **tutela dei diritti dei soggetti deboli**, tra cui minori, anziani, persone malate, disabili e non autosufficienti, incapaci, in taluni casi, di giungere a comprendere appieno e in modo autonomo le tecnologie utilizzate. Secondo lo Statuto di Fondazione Leonardo-Civiltà delle Macchine, chi sviluppa sistemi IA dovrebbe, in primis, pensare a tali persone e a come agevolarle nell'approcciarsi, in generale, alle nuove tecnologie.



È etica l'Intelligenza Artificiale che fa del bene all'intero ecosistema

L'emergenza pandemica ha investito anche il dibattito sull'etica dell'Intelligenza Artificiale che, negli ultimi due anni - complice il particolare momento storico, dominato da una maggiore consapevolezza circa i pericoli che corrono l'umanità e, con essa, tutto il pianeta - ha visto mutare la visione che si ha della tecnologia. Come ha spiegato Emanuela Girardi, membro del Direttivo AIxIA (Associazione Italiana per l'Intelligenza Artificiale), si sta verificando il passaggio «dalla visione europea antropocentrica in merito all'utilizzo dei sistemi AI a quella che pone al centro l'ecosistema, di cui l'uomo non è che una parte».

L'Intelligenza Artificiale ha ampliato il suo raggio di azione, ponendosi al servizio di cause dal respiro universale, come i 17 Obiettivi dell'Agenda 2030, inerenti - tra le tante tematiche a livello globale - anche alle questioni relative allo sviluppo economico e sociale, al cambiamento climatico e alla tutela dell'ambiente. Ecco, allora, che parlare di etica dell'AI, in questo momento, riferita a progetti di tale natura e rilevanza, significa porsi una serie di domande sulle finalità alle quali tali progetti rispondono e sulle eventuali problematiche nel conseguirli.



Ci si interroga sul “come” vengono realizzati i progetti AI finalizzati al bene sociale

Uno studio condotto dall’Oxford Internet Institute dell’Università di Oxford - pubblicato il 17 febbraio 2021 su Nature Machine Intelligence - indaga per la prima volta sulle “prove empiriche” dei progetti socialmente utili che si avvalgono dell’utilizzo delle tecniche di intelligenza artificiale, partendo dal presupposto che le “AI for social good” sono ormai in forte ascesa. E, nel farlo, prende come riferimento proprio gli Obiettivi di sviluppo sostenibile definiti dall’ONU, i Sustainable Development Goals (SDGs). Più nel dettaglio, il team di studio presenta una raccolta di iniziative di questo tipo, esaminando come vengono affrontate e portate a termine, tenendo conto anche delle criticità incontrate lungo il percorso.

Ricordiamo che tale studio rientra nell’Oxford Initiative on AI×SDGs, il cui fine è approfondire “come”, nel concreto, l’Intelligenza Artificiale è utile alla realizzazione dei 17 Obiettivi ONU.

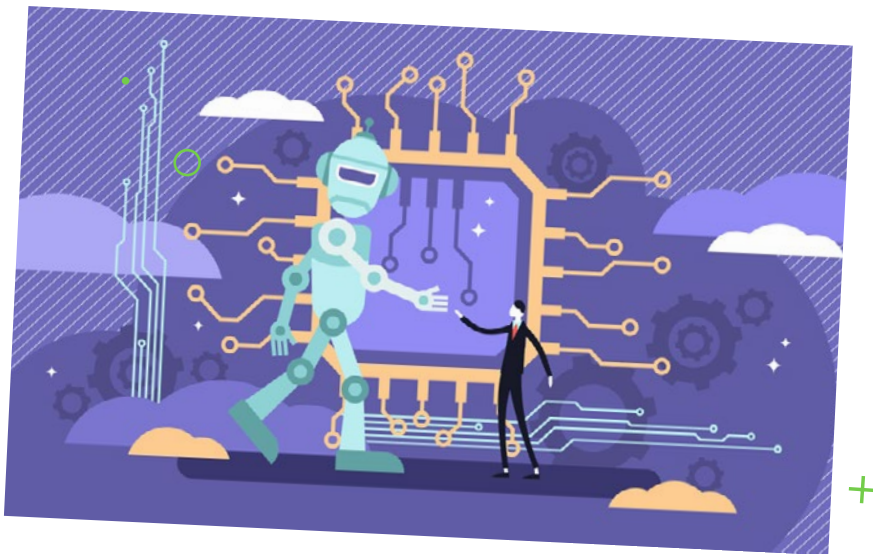


Lo scopo è definire uno strumento che, per coloro che operano nell’ambito dello sviluppo sostenibile, possa fungere da «linee guida etiche basate sull’evidenza, per futuri progetti di Intelligenza Artificiale finalizzati al bene sociale».



Le “linee guida etiche basate sull’evidenza”, ovvero le prove empiriche alle quali si è accennato, poggiano su alcune domande chiave, che gli eticisti si pongono nel momento in cui si interrogano sull’AI for social good. Tra queste, il riuscire a spiegare che cosa davvero conduce alla piena realizzazione un progetto che sfrutta l’Artificial Intelligence per il conseguimento di un obiettivo SDG e le eventuali differenze (e come possono essere superate e risolte) tra diverse aree geografiche, che spesso si accompagnano al raggiungimento di un dato obiettivo.

Un’altra domanda importante riguarda la possibilità di definire un metodo che possa facilitare futuri progetti AI sulla sostenibilità ambientale e di **mettere a punto un “codice etico di pratica” per supportare l’uso dell’Intelligenza Artificiale nell’affrontare gli obiettivi dell’Agenda 2030.**





L'etica degli algoritmi e l'attenzione alla loro trasparenza

Un altro studio a cura dell'Oxford Institute - pubblicato il 20 febbraio 2021 su AI & Society - si focalizza, invece, sul rischio etico legato agli algoritmi AI sviluppati col preciso obiettivo di

«migliorare il benessere individuale e sociale»,

ricordandoci che

«gli algoritmi non sono eticamente neutri».

Basti pensare - citano gli autori - agli algoritmi predittivi messi a punto per la gestione dei dati sanitari negli Stati Uniti, rei di "decidere", per i pazienti bianchi, cure considerate più efficienti rispetto a quelle somministrate ai pazienti di colore.

Riguardo, in particolare, al concetto di trasparenza riferito agli algoritmi AI (che devono poter essere sempre "spiegabili"), il team di studio sottolinea come la sua mancanza si traduca spesso in

«un'assenza di controllo, a sua volta responsabile di una mancanza di affidabilità».

Tra le cause della mancanza di trasparenza figurano, in molti casi, la difficoltà nell'analizzare e nell'interpretare vaste mole di dati, l'assenza di adeguati strumenti a disposizione e l'aver sotto mano dati di difficile lettura perché completamente destrutturati.

Sono, nello specifico, gli algoritmi di autoapprendimento quelli in cui la mancanza di trasparenza risulta essere una peculiarità propria, in quanto in grado di produrre nuove re-

gole nel momento stesso in cui apprendono. Il che contribuisce ad aumentare la difficoltà - per chi lavora al loro sviluppo - nel comprendere il passaggio dall'input all'output e il motivo stesso delle modifiche apportate.

Ebbene l'AI for social good - tra cui gli obiettivi di sostenibilità ambientale e sociale - deve tenere conto di tale rischio, analizzandolo caso per caso. Proprio dal metterlo a fuoco, dal comprenderlo, deriva una progettazione più trasparente degli algoritmi, per una governance dell'AI esplicitamente finalizzata al bene sociale.

+ **Quello dell'Intelligenza Artificiale per il bene comune è, per eccellenza, l'ambito in cui le due anime delle tecnologie digitali - il loro lato chiaro e il loro lato oscuro - sono chiamate a trovare il corretto equilibrio, l'armonia necessaria per "fare del bene".**





Riflessioni attorno alle **conseguenze** eticamente problematiche dell'Intelligenza Artificiale

Sono sufficienti le raccomandazioni e le linee guida definite da Organizzazioni mondiali, Governi e Istituzioni private a fare sì che vengano sviluppati sistemi di Intelligenza Artificiale etici e che questi vengano utilizzati in modo responsabile? Secondo Mariarosaria Taddeo, filosofa, esperta di etica delle tecnologie digitali e ricercatrice senior presso l'Internet Institute dell'Università di Oxford - la quale, a luglio del 2020, ha rilasciato un'intervista alla Redazione di Tech4Future - la questione è piuttosto complessa e richiede più livelli di analisi.



«C'è stata, negli ultimi dieci anni, un'attenzione crescente verso l'etica dell'Intelligenza Artificiale. Man mano che venivano sviluppati sempre nuovi algoritmi, e che venivano usati in maniera sempre più pervasiva, abbiamo iniziato a renderci conto delle loro conseguenze eticamente problematiche» spiega.

In primis, quello che è stato toccato con mano è il fatto che **molte delle decisioni che un sistema di Intelligenza Artificiale suggerisce rimangono non attribuibili in termini di responsabilità**. È difficile - osserva l'eticista - individuare chi è responsabile, sia moralmente sia legalmente, quando qualcosa "va storto" nell'utilizzo di un sistema IA.

E si è giunti, poi, alla conclusione che è un problema avere a che fare con una tecnologia in cui la tracciabilità dei sistemi e la disponibilità della documentazione relativa ai dati utilizzati, sono spesso impossibili.

«Ecco, man mano che tali questioni diventavano più pressanti, la domanda su come affrontarle è divenuta centrale, catalizzando anche l'interesse da parte di sviluppatori e ingegneri. Si è creata, dunque, una certa sensibilità attorno al tema. Ma la strada è ancora molto lunga».



L'attribuzione delle responsabilità e l'etica by Design

Per comprendere appieno la portata della riflessione, è importante chiarire un punto. Siamo abituati a pensare alla responsabilità morale come a quella responsabilità che si attribuisce a un agente che ha eseguito una data azione, fa notare Taddeo. Ma quando parliamo di Intelligenza Artificiale, parliamo anche di sistemi distribuiti, di più algoritmi sviluppati da aziende diverse e che poi vengono integrati:

«Poniamo l'esempio di un'auto a guida autonoma e ai numerosi algoritmi e alle tecnologie che la compongono. Se qualcosa non funzionasse, se accadesse un incidente e ci fossero una o più vittime, di chi sarebbe la colpa? Non ci sarebbe un solo attore responsabile, ma decine e decine. Il che rende estremamente complessa l'attribuzione di responsabilità. E al momento, non abbiamo ancora un modello concettuale che mappi il fenomeno».

Una seconda questione riguarda, invece, la possibile risposta ai problemi etici dell'IA proveniente dall'Etica by Design. Ma si tratta di una risposta sempre esaustiva?

Certo, sviluppare tecnologie che siano eticamente valide, è una buona soluzione. Ma rappresenta "una parte" della soluzione, non "la" soluzione, sottolinea l'esperta di etica.

Il fatto di implementare, già in fase di progettazione e sviluppo di un sistema di Intelligenza Artificiale, precise misure volte a risolvere in partenza i problemi etici che potrebbero derivare dal suo utilizzo - prosegue - ci permette di intervenire solo su questioni che già conosciamo.

L'Etica by Design, infatti, è una soluzione che non tiene conto di quelle che sono le conseguenze involontarie, non intenzionali, legate all'uso dell'IA.

Se i rischi connessi a un uso doloso dell'Intelligenza Artificiale derivano da un suo utilizzo intenzionalmente guidato da scopi illegali - tra cui, ad esempio, attacchi informatici, manipolazione dell'opinione pubblica attraverso fake news e campagne di disinformazione - i rischi connessi a un suo utilizzo non doloso, tra cui il pregiudizio non intenzionale e la discriminazione, proprio perché non controllabili, sono più insidiosi.

Si tratta, nello specifico, di rischi che derivano dalla natura stessa dell'IA,

«la prima tecnologia, nella storia dell'umanità, in grado di apprendere autonomamente. È capace di adattarsi perfettamente all'ambiente, all'interno del quale sviluppa quei comportamenti che rispondono meglio all'ecosistema in cui si trova. E il problema sta proprio qui: trattandosi di un agente autonomo che impara, non siamo in grado di prevederne i comportamenti involontari non etici».



I comportamenti dolosi (e involontari) dei sistemi IA

L'esempio più eclatante - ricorda Taddeo - è quello del software progettato da Microsoft per simulare conversazioni, lanciato su Twitter e su altre piattaforme social alcuni anni fa. Tay (questo il nome del chatbot) era stato programmato per imparare a chattare ripetendo frasi usate dagli utenti, per poi rispondere in modo autonomo, cercando di simulare una conversazione "normale".

Ebbene, il sistema, in poche ore, è diventato razzista, sessista e xenofobo, ma non perché il software fosse di per sé cattivo, ma semplicemente perché aveva raccolto dati e commenti su Twitter. E la maggior parte di quei dati e di quei commenti erano, evidentemente, razzisti, sessisti e xenofobi. Insomma, il sistema aveva rigurgitato quello che aveva imparato.

Un altro esempio, invece, è dato dall'utilizzo di una soluzione di Intelligenza Artificiale in alcuni Stati americani, per supportare le donne in gravidanza prossime al parto. Il sistema era stato creato e allenato su dati, prevalentemente, di donne caucasiche, le cui caratteristiche, dal punto di vista fisiologico, sono diverse rispetto a quelle delle donne afro-americane e ispaniche. Particolare, questo, che avrebbe potuto creare problemi a queste ultime, in quanto i loro profili clinici non erano in linea con il sistema. **Anche in questo caso, da una buona intenzione di partenza, si è passati a un risultato decisamente negativo, con il rischio di un danno grave sotto il profilo psico-fisico dei soggetti coinvolti.**



Come mitigare i rischi di un'Intelligenza Artificiale non etica

Mariarosaria Taddeo, alla domanda sulle possibili, concrete contromisure da adottare per fare fronte alle conseguenze eticamente problematiche dell'Intelligenza Artificiale, risponde facendo riferimento a "un insieme" di strumenti.

L'approccio etico by design, certo, è utile, ma occorre - secondo la filosofa - anche un training per sviluppatori e ingegneri sulle questioni etiche legate all'IA, oltre all'acquisizione di una maggiore consapevolezza da parte degli utenti. E poi c'è un terzo elemento, ovvero il monitoraggio, il controllo costante di queste tecnologie, per capire quando stanno deragando rispetto ai valori che vedono al centro la persona.

Non è solo un problema di regolamentazioni che vengono dall'alto - precisa - di linee guida definite dai Governi, né di questioni che riguardano soltanto coloro che sviluppano le IA.

«È una questione che richiede un multi-stakeholders approach, a partire da provider, aziende, utenti e Governi. Tutti hanno un ruolo e tutti hanno delle responsabilità».

È il momento di passare dai principi alle **azioni concrete**

«In tema di Intelligenza Artificiale ed etica, la parola d'ordine, ora, è "concretezza". Sulla materia è stata prodotta letteratura in abbondanza. È importante che adesso si passi dai principi alle azioni concrete». È il pensiero di Francesca Rossi, AI Ethics Global Leader IBM - intervistata dalla Redazione di Tech4Future a settembre del 2020 - la quale sottolinea l'importanza, all'interno delle aziende IA, di dedicare tempo e sforzi a quelle attività che aiutino tutte le divisioni a capire che cosa davvero significhi, nell'ambito del lavoro quotidiano, rispettare i requisiti etici dell'Intelligenza Artificiale.



Il dibattito tra tecnici IA ed eticisti

Convinta che quello tra tecnici IA ed esperti di etica non sia un dibattito tra due gruppi separati, distanti tra loro, ma tra due sfere che si integrano, Francesca Rossi fa notare come, all'interno delle Università di tutto il mondo, durante le lezioni dei corsi di informatica, di parla anche di aspetti etici.

E sono sempre più numerosi - spiega - i tecnici che conoscono gli aspetti dell'etica e gli eticisti che, a loro volta, possiedono un sapere tecnico in fatto di artificial intelligence. È vero che, per molto tempo, i ricercatori sono stati un po' restii ad aprirsi a questo genere di dibattito ma, da almeno sei anni-sette anni le cose sono cambiate e oggi c'è più apertura al tema.

E chi afferma che "non esiste un'etica dell'Intelligenza Artificiale", in realtà si riferisce al fatto che esiste un certo conflitto, una certa tensione, tra quella che è l'attenzione ai problemi etici dell'IA da un lato e la business value, il profitto che proviene dall'IA, dall'altro.

«Personalmente, penso che questa tensione sia solo apparente e che i due ambiti - etica e profitto dell'Artificial Intelligence - vadano di pari passo. Anzi, il modo ottimale per migliorare, affinare, le capacità dell'AI e, allo stesso tempo,

per trarne profitto, è includere nel suo studio e nel suo sviluppo anche i principi etici.

I clienti stessi - ad esempio - ci chiedono in che modo, in IBM, affrontiamo l'etica rispetto ai sistemi di Intelligenza Artificiale. La sensibilità nei confronti di questo tema sta sempre più diventando motivo di differenziazione per le aziende del settore».



Cosa significa ragionare in termini di “etica” quando si parla di Intelligenza Artificiale

Trattare di etica dell'IA significa, innanzitutto, guardare in faccia le preoccupazioni, i timori che suscita in noi l'insieme delle tecnologie che fanno capo all'ambito di studi dell'Artificial Intelligence.

Timori e preoccupazione che, in realtà - osserva Rossi - sorgono di fronte a qualsiasi tecnologia pervasiva, in grado di portare trasformazioni nella nostra vita e il cui impatto può, talvolta, spaventarci.

Nel caso specifico dell'Intelligenza Artificiale, i timori sono legati proprio al tipo di tecnologia. Ad esempio, al fatto che, tra le tecniche che utilizza, ve ne sono alcune che appaiono come “scatole nere”, il cui meccanismo, proprio come una black box, è descrivibile solo guardando al comportamento esterno, in quanto riamane incomprensibile come, da deter-

minati dati in ingresso, derivi, poi, una determinata decisione in uscita. Cosa c'è dentro la scatola nera resta ignoto ai non addetti ai lavori.

In particolare, il riferimento è al machine learning: che cosa preoccupa, nel concreto, di questa tecnologia?

«Il fatto di non riuscire a capire in che modo - e perché - prenda determinate decisioni, rende insicuri e fa sì che non ci si affidi tranquillamente a una macchina del genere, soprattutto nel seguire le raccomandazioni e i suggerimenti che questa dà a chi la utilizza e la interroga. Questo è un esempio tipico di "paura" generata dall'AI».

Un altro esempio ancora - prosegue - è correlato al suo funzionamento, che poggia su una grande quantità di dati. Da qui scaturiscono tutta una serie di timori sulla loro gestione, su chi c'è dietro questi dati, sul loro reale utilizzo. E poi, sempre riguardo alle decisioni che prendono e ai suggerimenti che danno, la domanda frequente che ci si pone è se queste macchine sappiano distinguere tra una decisione buona e una cattiva, se sono attente a non prendere decisioni che potrebbero essere discriminatorie.





Perché una macchina dovrebbe prendere decisioni discriminatorie?

Chi sviluppa sistemi AI - rimarca l'AI Ethics Global Leader IBM - non trasferisce loro tutti i passaggi da compiere per risolvere un determinato problema e prendere la decisione giusta. Alle macchine non vengono date soluzioni, bensì esempi.

Questo significa che è di fondamentale importanza inserire nella macchina esempi corretti dal punto di vista etico: se vogliamo che un sistema sia in grado di discernere tra una richiesta di mutuo con i requisiti giusti per poter essere accettata e una proposta che, invece, non li ha, e lo alleniamo dandogli esempi diversi di richieste, specificando, per ognuna, se sia idonea, dobbiamo stare attenti a non trasferire modelli in cui - ad esempio - tutte le richieste di mutuo da parte di persone giovani vengono accettate, mentre quelle che provengono da parte di persone anziane vengono rifiutate. Dandole - non intenzionalmente - la correlazione tra età degli utenti e opzioni di rifiuto o di accettazione, mettiamo la macchina nelle condizioni di operare scelte discriminatorie nei confronti di un target di età specifico.

Ed è qui che entrano in gioco gli strumenti regolatori messi a punto a livello internazionale e nazionale, aiutando gli sviluppatori

«a divenire più consapevoli circa i requisiti dell'Intelligenza Artificiale etica, a seguirli, a inserirli nei loro processi di lavoro, ad applicarli nel momento in cui danno vita a un nuovo algoritmo o a un nuovo sistema di artificial intelligence».

Conclusioni

di Paola Cozzi

Se provassimo a sfrondare il dibattito sull'etica dell'Intelligenza Artificiale per andare dritti al suo nocciolo, ci accorgemmo che stiamo parlando di noi stessi. Il comportamento dell'IA è il nostro comportamento. L'etica dell'IA è l'etica dell'essere umano.

Indichiamo l'algoritmo, la macchina, ma a progettarli, a svilupparli e ad addestrarli siamo noi. Dovremmo indicare noi stessi. E non tanto per un'attribuzione di meriti o di colpe, per ripeterci quanto siamo stati "bravi" o "cattivi" a concepirli tanto potenti, quanto per darci delle responsabilità.

La riflessione sul "darsi delle responsabilità" porta ad annullare la contrapposizione dialettica tra negativo-positivo (l'Intelligenza Artificiale è cattiva? È buona?), per enfatizzare, invece, i concetti di "mezzo" e di "obiettivo". Come sostiene Olle Häggström, docente di Statistica Matematica alla Chalmers University di Göteborg, in Svezia, nel documento "Should we fear artificial intelligence?", a cura dello European Parliamentary Research Service, «i sistemi IA sono solo strumenti. Essi ci consentono di raggiungere gli obiettivi che sapremo affidare loro. E se tali obiettivi saranno contrari o a favore dell'umanità, dipenderà unicamente da noi».

Attribuire a noi stessi la responsabilità del comportamento delle macchine dotate di Intelligenza Artificiale non semplifica, certo, le spinose questioni sul tavolo, ma è utile a smontare le preoccupazioni e le paure per i suoi possibili sviluppi futuri malevoli ai danni della persona e a riportare il discorso su "che cosa" la persona si aspetta dalla macchina.

In un passaggio dell'intervista all'eticista Mariarosaria Taddeo, viene sottolineato il fatto che l'etica dell'IA non è solo un problema di "regole" dall'alto, né di nodi che riguardano soltanto coloro che le sviluppano, ma è un problema di tutti,

«a partire da provider, aziende, utenti e Governi. Tutti hanno un ruolo e tutti hanno delle responsabilità».

La responsabilità che abbiamo nei confronti dell'Intelligenza Artificiale dovrebbe indurci a modulare le nostre scelte in merito al suo utilizzo in determinati ambiti, soprattutto laddove la sua applicazione potrebbe avere un impatto sulle decisioni dei soggetti (ad esempio, nella gestione del personale) o sugli aspetti psicologici ed emotivi dell'individuo, come, ad esempio, in ambito educativo.

Se l'etica dell'IA è un problema di tutti e tutti noi abbiamo un ruolo affinché il suo utilizzo sia eticamente corretto, le domande che tutti dovremmo imparare a porci sono: **“ho davvero bisogno dell'IA in questo contesto? Farà realmente del bene, in questo ambito, il suo utilizzo?” “Che rischi corriamo se la adottiamo in riferimento a questa specifica situazione?”**.

È l'analisi dei rischi - di cui si legge, peraltro, nella proposta di Regolamento della Commissione europea del 21 aprile 2021 per un'IA sicura - il nuovo filo rosso che lega tutti i punti della discussione attorno all'etica dell'Intelligenza Artificiale e che, ci auguriamo, possa tracciare, in futuro, la nostra rotta di comportamento. Comportamento dal quale dipenderà il comportamento dell'IA.

TECH4FUTURE

www.tech4future.info

Direttrice responsabile
Nicoletta Boldrini

Caporedattrice
Paola Cozzi

Progetto grafico e impaginazione
Lucia Cariani