< >

Montréal Declaration
Responsible AI_

</ >

DICHIARAZIONE DI MONTRÉAL SULLO SVILUPPO RESPONSABILE DELL'INTELLIGENZA ARTIFICIALE

2018

Il presente documento fa parte della **DICHIARAZIONE DI MONTRÉAL PER LO SVILUPPO RESPONSABILE DELL'INTELLIGENZA ARTIFICIALE** 2018. Il rapporto completo è disponibile QUI.

INDICE

LETTURA DELLA DICHIARAZIONE	5
PREMESSA	7
PRINCIPI	
1. PRINCIPIO DI BENESSERE	8
2. PRINCIPIO DI RISPETTO PER L'AUTONOMIA	9
3. PROTEZIONE DI PRIVACY E RISERVATEZZA	10
4. PRINCIPIO DI SOLIDARIETÀ	11
5. PRINCIPIO DI PARTECIPAZIONE DEMOCRATICA	12
6. PRINCIPIO DI EQUITÀ	13
7. PRINCIPIO DI INCLUSIONE DELLE DIVERSITÀ	14
8. PRINCIPIO DI CAUTELA	15
9. PRINCIPIO DI RESPONSABILITÀ	16
10. PRINCIPIO DI SVILUPPO SOSTENIBILE	17
GLOSSARIO	18
CREDITI	I
PARTNER	II

LETTURA DELLA DICHIARAZIONE

PERCHÉ UNA DICHIARAZIONE?

La Dichiarazione di Montréal per uno sviluppo responsabile dell'IA ha tre obiettivi primari:

1. Sviluppare un quadro etico per lo sviluppo e l'implementazione dell'IA;
2. Guidare la transizione digitale in modo che tutti traggano vantaggio dalla rivoluzione tecnologica;
3. Avviare un forum di discussione a livello nazionale e internazionale con l'obiettivo di sviluppare l'IA in modo equo, inclusivo ed ecosostenibile.

DICHIARARE COSA?

I PRINCIPI

Il primo obiettivo della Dichiarazione è quello di individuare i principi e i valori etici che favoriscono gli interessi fondamentali delle persone e dei gruppi. Tali principi applicati al campo dell'intelligenza artificiale rimangono generali e astratti. Per leggerli correttamente, è importante tenere a mente i punti seguenti:

> Sebbene vengano presentati sotto forma di elenco, non sono in ordine di importanza. L'ultimo principio non è meno importante del primo. Tuttavia, a seconda delle circostanze è possibile che venga dato più peso a un principio piuttosto che a un altro, o che si consideri più rilevante un principio rispetto a un altro.

> Sebbene siano diversi, devono essere interpretati in modo coerente, al fine di evitare conflitti che ne impediscano l'applicazione. In linea di massima, i limiti di applicazione di un principio vengono delineati dal campo di applicazione di un altro principio.

> Sebbene riflettano la morale e la cultura politica della società in cui sono stati sviluppati, gettano le basi per un dialogo interculturale e internazionale.

> Sebbene possano essere interpretati in modi diversi, non possono avere un'interpretazione qualsiasi. È imperativo che l'interpretazione sia coerente.

> Sebbene si tratti di principi etici, possono essere tradotti nel linguaggio politico e interpretati da un punto di vista legale.

Sulla base di detti principi sono state espresse raccomandazioni, al fine di stabilire delle linee guida per la transizione digitale nell'ambito del quadro etico della Dichiarazione. L'obiettivo è trattare alcuni temi trasversali da inserire nella transizione verso una società in cui l'IA possa contribuire al bene comune: gestione degli algoritmi, alfabetizzazione digitale, inclusione digitale delle diversità ed ecosostenibilità.

DICHIARAZIONE PER CHI?

La Dichiarazione di Montréal si rivolge a tutti gli individui, le organizzazioni e le aziende che desiderano far parte di uno sviluppo responsabile dell'intelligenza artificiale, fornendo contributi scientifici o tecnologici, sviluppando progetti sociali, elaborando regole (normative e codici) a cui assoggettarla, per contrastare approcci malevoli o avventati, ovvero per allertare l'opinione pubblica, se necessario.

Si rivolge inoltre ai rappresentanti politici, eletti o nominati, che i cittadini auspicano si occupino di sviluppare cambiamenti sociali, introdurre rapidamente un quadro di attuazione della transizione digitale al servizio del bene comune e prevedano i gravi rischi posti dallo sviluppo dell'IA.

UNA DICHIARAZIONE SECONDO QUALE METODO?

La Dichiarazione è nata da un processo di discussione inclusiva che ha attivato il dialogo tra cittadini, esperti, pubblici ufficiali, rappresentanti del settore, organizzazioni civili e associazioni professionali. I vantaggi offerti da questo approccio sono triplici:

1. Mediare collettivamente le controversie sociali ed etiche dell'IA;
2. Migliorare la qualità della riflessione sull'IA responsabile;
3. Consolidare la legittimità delle proposte per l'IA responsabile.

L'elaborazione di principi e raccomandazioni è un lavoro di co-costruzione che ha coinvolto una vasta gamma di partecipanti negli spazi pubblici, nelle sale riunioni delle organizzazioni professionali, nelle tavole rotonde degli esperti internazionali, nei laboratori di ricerca, nelle aule e online, sempre con lo stesso rigore.

DOPO LA DICHIARAZIONE?

Poiché si tratta di una tecnologia in continua crescita dagli anni '50 del '900 con una velocità di innovazione esponenziale, è fondamentale considerare la Dichiarazione come un documento guida aperto a revisioni e modifiche in base agli sviluppi delle conoscenze e delle tecnologie, nonché del feedback degli utenti sull'utilizzo dell'IA nella società. Al termine del processo di elaborazione della Dichiarazione, abbiamo raggiunto il punto di partenza di una discussione aperta e inclusiva sul futuro dell'umanità servita dalle tecnologie dell'intelligenza artificiale.

PREMESSA

Per la prima volta nella storia del genere umano, possiamo creare sistemi autonomi in grado di eseguire compiti complessi finora riservati all'intelligenza naturale: elaborare grandi quantità di informazioni, eseguire calcoli e previsioni, apprendere e adeguare le risposte alle situazioni mutevoli, riconoscere e classificare gli oggetti. Data la natura immateriale di questi compiti, e per analogia con l'intelligenza umana, definiamo questi sistemi ad ampio raggio con il termine generico di intelligenza artificiale. L'intelligenza artificiale rappresenta una forma elevata di progresso scientifico e tecnologico, in grado di generare considerevoli vantaggi sociali, migliorando le condizioni di vita e di salute, agevolando la giustizia, creando ricchezza e mitigando l'impatto delle attività umane sull'ambiente e sul clima. Le macchine intelligenti non si limitano a eseguire i calcoli meglio degli esseri umani, ma sono anche in grado di interagire con esseri senzienti, tener loro compagnia e prendersene cura.

Tuttavia, lo sviluppo dell'intelligenza artificiale presenta quesiti etici e rischi sociali. Praticamente, le macchine intelligenti possono limitare la scelta degli individui e dei gruppi di persone, ridurre lo standard di vita, turbare l'organizzazione e il mercato del lavoro, influenzare la politica, ledere i diritti fondamentali, esacerbare le diseguaglianze socio-economiche e modificare l'ambiente. Sebbene il progresso scientifico e la vita in società comportino sempre dei rischi, è diritto dei cittadini stabilire le finalità morali e politiche che danno senso ai rischi posti da un determinato mondo.

Più basso è il rischio di implementazione, maggiori sono i vantaggi offerti dell'intelligenza artificiale. Il primo rischio posto dallo sviluppo dell'intelligenza artificiale è il fatto di dare l'impressione di poter controllare il futuro con i calcoli. Ridurre la società a una serie di numeri e governarla tramite algoritmi, è un sogno antico che tuttora spinge l'ambizione umana. Ma quando si tratta di questioni umane, raramente il domani somiglia a oggi e i numeri non possono capire cosa possiede valore morale o è socialmente auspicabile.

I principi di questa Dichiarazione sono come i punti di una bussola morale, che guidano lo sviluppo dell'intelligenza artificiale verso obiettivi morali e sociali auspicabili. Offrono inoltre un quadro etico che promuove i diritti umani riconosciuti a livello internazionale nei campi coinvolti nell'applicazione dell'intelligenza artificiale. Complessivamente, i principi enunciati gettano le basi per una crescita della fiducia sociale nei sistemi di intelligenza artificiale.

I principi di questa Dichiarazione sono basati sulla convinzione comune che gli esseri umani cerchino di crescere come esseri sociali dotati di sensazioni, pensieri e sentimenti, e ambiscano a realizzare il proprio potenziale esercitando liberamente le proprie capacità emotive, morali e intellettive. È compito dei vari soggetti interessati pubblici e privati e decisori politici a livello locale, nazionale e internazionale, fare in modo che lo sviluppo e l'applicazione dell'intelligenza artificiale siano compatibili con la difesa delle capacità e degli obiettivi umani fondamentali, e contribuiscano alla loro piena realizzazione. Avendo chiaro questo obiettivo, bisogna interpretare i principi proposti in maniera coerente, tenendo presente il contesto socio-politico-culturale e legale specifico della loro applicazione.

1

PRINCIPIO DI **BENESSERE**

Lo sviluppo e l'uso dei sistemi di intelligenza artificiale (IA) devono consentire il miglioramento del benessere di tutti gli esseri senzienti.

1. I sistemi IA devono aiutare le persone a migliorare le proprie condizioni di vita, di salute e di lavoro.

2. I sistemi IA devono permettere alle persone di perseguire le proprie attitudini, nella misura in cui non ledano altri esseri senzienti.

3. I sistemi IA devono permettere alle persone di mettere in pratica le proprie capacità mentali e fisiche.

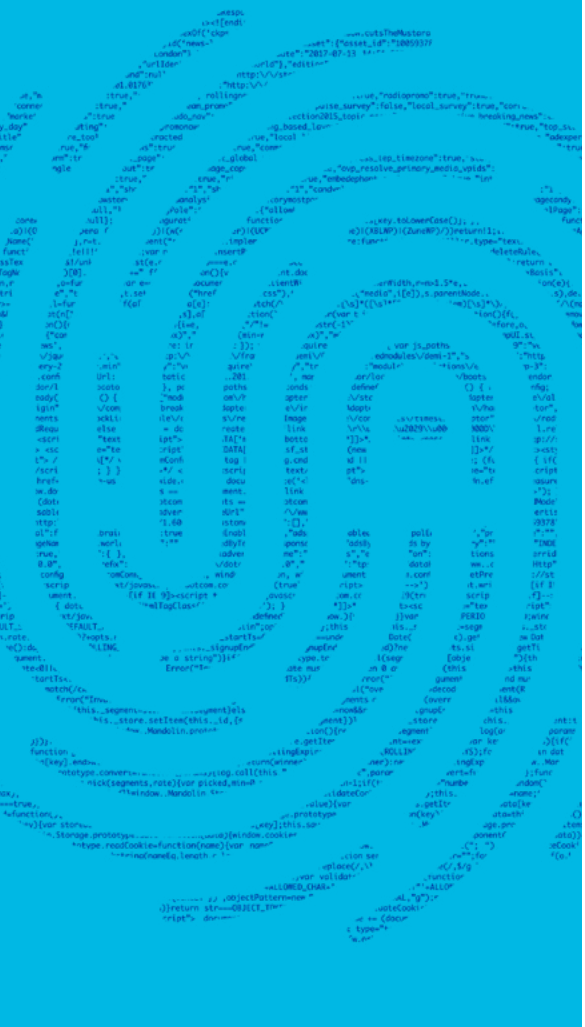
4. I sistemi IA non devono diventare una fonte di malessere, a meno che questo non preluda al conseguimento di un benessere superiore a quanto si otterrebbe altrimenti.

5. L'utilizzo dei sistemi IA non deve aumentare stress, ansia e la sensazione di essere vessati dall'ambiente digitale.

2

PRINCIPIO DI **RISPETTO PER L'AUTONOMIA**

I sistemi AIS devono essere sviluppati e utilizzati nel rispetto dell'autonomia delle persone e con l'obiettivo di aumentarne il controllo sulla propria vita e sull'ambiente circostante.



1. I sistemi IA devono permettere alle persone di realizzare i propri obiettivi e la propria concezione di vita degna di essere vissuta.

2. I sistemi IA non devono essere sviluppati e utilizzati per imporre alle persone un particolare stile di vita, né direttamente né indirettamente, applicando metodi di sorveglianza oppressiva o meccanismi di valutazione o incentivazione.

3. Le pubbliche istituzioni non devono utilizzare i sistemi IA per promuovere o ostacolare una particolare idea di vita buona.

4. È fondamentale abituare i cittadini alle tecnologie digitali garantendo l'accesso alle relative forme di conoscenza, favorendo l'apprendimento delle competenze fondamentali (alfabetizzazione digitale e multimediale) e sostenendo lo sviluppo del pensiero critico.

5. I sistemi IA non devono essere sviluppati per diffondere informazioni inattendibili, menzognere o di propaganda, e devono essere progettati con l'obiettivo di limitarne la diffusione.

6. Lo sviluppo dei sistemi IA non deve creare dipendenza con tecniche capaci di calamitare l'attenzione o imitando le caratteristiche umane (aspetto, voce, ecc.) in un modo che possa causare confusione tra IA ed esseri umani.

3

PRINCIPIO DI **PROTEZIONE DELLA PRIVACY E DELLA RISERVATEZZA**

Privacy e riservatezza devono essere protette dall'intromissione dei sistemi IA e dei sistemi di acquisizione e archiviazione dei dati (AAD).

1. Gli spazi personali in cui le persone non sono soggette a sorveglianza o a valutazione digitale devono essere protetti dall'intromissione dei sistemi IA e AAD.
2. La riservatezza di pensieri ed emozioni deve essere rigorosamente protetta dall'utilizzo di sistemi IA e AAD che possano recare danno, in particolare imponendo giudizi morali sulle persona e le relative scelte di vita.
3. Gli individui devono sempre mantenere il diritto a scollegare la propria vita privata dai sistemi digitali e i sistemi IA devono offrire esplicitamente l'opzione di disconnessione a intervalli regolari, senza incoraggiare le persone a restare connesse.
4. Le persone devono mantenere il controllo completo sulle informazioni relative alla proprie preferenze. I sistemi IA non devono creare profili delle preferenze individuali per influenzare il comportamento delle persone, senza il loro consenso libero e informato.
5. I sistemi AAD devono garantire la riservatezza dei dati e l'anonimato del profilo personale.
6. Ognuno deve poter esercitare pieno controllo sui propri dati personali, soprattutto in relazione alla loro raccolta, utilizzo e divulgazione. L'accesso delle persone ai sistemi IA e ai servizi digitali non deve essere vincolato alla rinuncia al controllo o alla proprietà dei dati personali.
7. Le persone devono essere libere di fornire i propri dati alle organizzazioni di ricerca per contribuire al progresso delle conoscenze.
8. Deve essere garantita l'integrità della propria identità personale. I sistemi IA non devono essere utilizzati per imitare o alterare l'aspetto, la voce o altre caratteristiche individuali dell'utente, al fine di danneggiarne la reputazione o manipolare altre persone.

4

PRINCIPIO DI **SOLIDARIETÀ**

Lo sviluppo dei sistemi IA deve essere compatibile con il mantenimento dei legami di solidarietà tra persone e generazioni.

1. I sistemi IA non devono minacciare il mantenimento delle relazioni umane morali ed emotive, mentre devono essere sviluppati con l'obiettivo di favorire tali relazioni e ridurre la vulnerabilità e l'isolamento delle persone.

2. I sistemi IA devono essere sviluppati con l'obiettivo di aiutare gli esseri umani a svolgere compiti complessi e devono favorire la collaborazione tra esseri umani.

3. I sistemi IA non devono essere utilizzati per sostituire le persone in ambiti che richiedano relazioni umane di qualità, ma devono invece mirare ad agevolare tali relazioni.

4. I sistemi sanitari che utilizzano l'IA devono tenere conto dell'importanza delle relazioni del paziente con i familiari e il personale sanitario.

5. Lo sviluppo dei sistemi IA non deve incoraggiare comportamenti ostili verso i robot destinati a somigliare a esseri umani o animali nell'aspetto o nei comportamenti.

6. I sistemi IA devono contribuire a migliorare la gestione dei rischi e facilitare le condizioni per una società con una più equa e comune distribuzione dei rischi individuali e collettivi.

5

PRINCIPIO DI **PARTECIPAZIONE DEMOCRATICA**

I sistemi IA devono soddisfare i criteri di intelligibilità, giustificabilità e accessibilità, e devono essere soggetti ad attento esame, dibattito e controllo democratico.

1. I processi IA che incidono sulla vita delle persone, sulla qualità della vita o sulla reputazione degli individui, devono essere chiari a chi li crea.

2. Le decisioni prese dai sistemi IA che incidono sulla vita delle persone, sulla qualità della vita o sulla reputazione degli individui, devono sempre essere giustificabili, con un linguaggio comprensibile per le persone che li usano o che subiscono le conseguenze del loro uso. La giustificazione consiste nel rendere trasparenti i principali fattori e i parametri che determinano le decisioni e deve assumere la stessa forma delle giustificazioni che sarebbero richieste a un essere umano che prenda lo stesso tipo di decisioni.

3. Il codice degli algoritmi pubblici e privati deve sempre essere accessibile alle autorità pubbliche competenti e ai soggetti interessati, per fini di verifica e controllo.

4. L'individuazione nei sistemi IA di errori operativi, effetti inattesi o indesiderati, violazioni della sicurezza e perdite di dati deve essere tassativamente segnalata alle autorità pubbliche competenti, ai soggetti interessati e a chiunque sia coinvolto nella condizione.

5. In conformità con il requisito di trasparenza delle decisioni pubbliche, il codice degli algoritmi decisionali utilizzati dalle autorità pubbliche deve essere accessibile a tutti, a eccezione degli algoritmi che presentino un elevato rischio di danni gravi in caso di uso improprio.

6. Per i sistemi IA che esercitano un impatto significativo sulla vita dei cittadini, devono essere garantite agli stessi la possibilità e le competenze per decidere i parametri sociali di tali sistemi, i loro obiettivi e limiti d'uso.

7. Dobbiamo sempre essere in grado di verificare che i sistemi IA eseguano ciò per cui sono stati programmati e per cui vengono utilizzati.

8. Chiunque utilizzi un servizio, deve sapere se una decisione che li riguarda o li condiziona sia stata presa da un sistema IA.

9. Chiunque utilizzi un servizio che prevede chatbot, deve essere in grado di individuare facilmente se l'interazione avvenga con un sistema IA o con una persona reale.

10. La ricerca sull'intelligenza artificiale deve restare aperta e accessibile a tutti.

6

PRINCIPIO DI **EQUITÀ**

Lo sviluppo e l'utilizzo dei sistemi IA devono contribuire alla creazione di una società giusta ed equa.

1. I sistemi IA devono essere progettati e addestrati in modo da non creare, rafforzare o produrre discriminazioni in base, tra l'altro, a differenze sociali, sessuali, etniche, culturali o religiose.

2. Lo sviluppo dei sistemi IA deve aiutare ad abolire le relazioni di dominio tra gruppi e persone sulla base di disparità di potere, ricchezza o conoscenza.

3. Lo sviluppo dei sistemi IA deve produrre vantaggi sociali ed economici per tutti, riducendo le disuguaglianze e le vulnerabilità sociali.

4. Lo sviluppo di sistemi IA industriali deve essere compatibile con condizioni di lavoro accettabili in ogni fase del loro ciclo di vita, dall'estrazione delle risorse naturali al riciclo, compreso il trattamento dei dati.

5. L'attività digitale degli utenti dei sistemi IA e dei servizi digitali deve essere riconosciuta come lavoro che contribuisce al funzionamento degli algoritmi e crea valore.

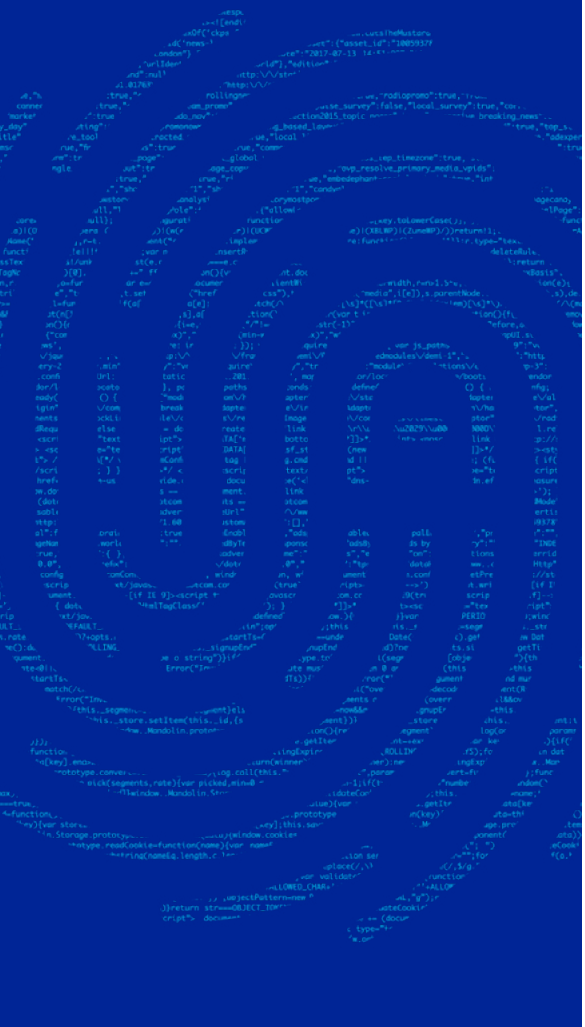
6. Deve essere garantito a tutti l'accesso a risorse, conoscenze e strumenti digitali fondamentali.

7. Dobbiamo sostenere lo sviluppo di algoritmi comuni, e dei dati aperti necessari per effettuarne il training, ed espanderne l'uso, come obiettivo socialmente equo.

7

PRINCIPIO DI **INCLUSIONE DELLE DIVERSITÀ**

Lo sviluppo e l'utilizzo dei sistemi IA deve essere compatibile con il mantenimento delle diversità socio-culturali e non devono limitare l'ambito delle scelte di vita o delle esperienze personali.



1. Lo sviluppo e l'utilizzo dei sistemi IA non deve portare all'omogeneizzazione sociale tramite la standardizzazione di comportamenti e opinioni.

2. Dal momento in cui gli algoritmi vengono ideati, lo sviluppo e l'applicazione dei sistemi IA deve prendere in considerazione la molteplicità di espressione delle diversità socio-culturali presenti nella società.

3. Gli ambienti di sviluppo dei sistemi IA, industriali o di ricerca, devono essere inclusivi e riflettere le diversità dei singoli individui e gruppi della società.

4. I sistemi IA non devono utilizzare i dati acquisiti per associare le persone a un profilo utente, stabilirne l'identità personale o isolarle in una bolla di filtraggio, in modo da limitarne e restringerne le possibilità di sviluppo personale, soprattutto negli ambiti della formazione, della giustizia o del business.

5. I sistemi IA non devono essere sviluppati o utilizzati per limitare la libera espressione delle idee o la possibilità di sentire opinioni diverse, entrambe condizioni essenziali per una società democratica.

6. Per ogni categoria di servizi, l'offerta di sistemi IA deve essere diversificata, in modo da evitare di fatto la formazione di monopoli e la riduzione delle libertà individuali.

8

PRINCIPIO DI **PRUDENZA**

Chiunque sia coinvolto nello sviluppo dei sistemi IA deve prestare attenzione a prevedere nei limiti del possibile le conseguenze negative dell'utilizzo dei sistemi IA e ad adottare misure adeguate a evitarle.

1. È necessario sviluppare meccanismi che tengano in considerazione il duplice utilizzo (benefico e dannoso) della ricerca e dello sviluppo dei sistemi IA (pubblici e privati) al fine di limitarne l'uso nocivo.

2. Se l'utilizzo improprio dei sistemi IA danneggia la salute pubblica o la sicurezza ed è altamente probabile che questo avvenga, è prudente limitare l'accesso libero e la divulgazione pubblica degli algoritmi.

3. Prima dell'immissione sul mercato e sia che vengano proposti a titolo gratuito o a pagamento, i sistemi IA devono soddisfare stringenti requisti di affidabilità, sicurezza e integrità, ed essere soggetti a test che non mettano in pericolo la vita delle persone, non ne danneggino la qualità della vita, la reputazione o l'integrità psicologica. Questi test devono essere aperti alle autorità pubbliche competenti e ai soggetti interessati.

4. Lo sviluppo dei sistemi IA devono prevenire i rischi di utilizzo improprio dei dati degli utenti e proteggere l'integrità e la riservatezza delle informazioni personali.

5. Errori e difetti individuati nei sistemi IA e AAD devono essere divulgati pubblicamente su scala globale dalle istituzioni pubbliche e dalle aziende dei settori che danneggiano in modo rilevante l'integrità personale e la struttura sociale.

9

PRINCIPIO DI **RESPONSIBILITÀ**

Lo sviluppo e l'utilizzo dei sistemi IA non devono contribuire a ridurre la responsabilità degli esseri umani di fronte alla necessità di assumere decisioni.

1. Solo gli esseri umani possono essere considerati responsabili delle decisioni derivanti dalle raccomandazioni dei sistemi IA e dalle azioni che le precedono.

2. In tutti gli ambiti in cui devono essere assunte decisioni che incidono sulla vita delle persone, sulla qualità della vita o sulla reputazione degli individui, ove i tempi e le circostanze lo consentano, la decisione definitiva deve essere libera e informata, e presa da un essere umano.

3. La decisione di uccidere deve sempre essere presa dagli esseri umani, senza trasferirne la responsabilità ai sistemi IA.

4. Chiunque autorizzi i sistemi IA a commettere un crimine o un reato, oppure dimostri omissione consentendo ai sistemi IA di compierli, è responsabile di tale crimine o reato.

5. Se sono stati prodotti danni da un sistema IA ritenuto affidabile utilizzato secondo gli usi previsti, non è ragionevole attribuire la colpa alle persone coinvolte nel suo sviluppo o utilizzo.

10

PRINCIPIO DI **SVILUPPO SOSTENIBILE**

Lo sviluppo e l'utilizzo dei sistemi IA devono essere effettuati in modo da garantire una forte sostenibilità ambientale del pianeta.

1. L'hardware dei sistemi IA, l'infrastruttura digitale e i relativi oggetti su cui si basa, come i centri dati, devono ambire alla massima efficienza energetica e alla riduzione delle emissioni di gas serra durante l'intero ciclo di vita.

2. L'hardware dei sistemi IA, l'infrastruttura digitale e i relativi oggetti su cui si basa, come i centri dati, devono ambire a generare la minore quantità possibile di rifiuti elettrici ed elettronici, e a fornire procedure di manutenzione, riparazione e riciclo secondo i principi dell'economia circolare.

3. L'hardware dei sistemi IA, l'infrastruttura digitale e i relativi oggetti su cui si basa, come i centri dati, devono ridurre al minimo l'impatto umano sugli ecosistemi e la biodiversità in ogni fase del loro ciclo di vita, soprattutto in relazione al rispetto dell'estrazione delle risorse e dello smaltimento finale delle apparecchiature al termine del ciclo di vita utile.

4. Gli attori pubblici e privati sono tenuti a sostenere lo sviluppo ecosostenibile dei sistemi IA, al fine di contrastare lo spreco di risorse naturali e di merci prodotte, generare commercio e catene di fornitura sostenibili, e ridurre l'inquinamento globale.

GLOSSARIO

Affidabilità
I sistemi IA sono affidabili quando eseguono i compiti a cui sono destinati nel modo previsto. L'affidabilità è la probabilità di successo compresa tra il 51% e il 100%, quindi necessariamente superiore al caso. Più un sistema è affidabile, più ne sono prevedibili i comportamenti.

Alfabetizzazione digitale
L'alfabetizzazione digitale di una persona indica la capacità di accedere, gestire, comprendere, integrare, comunicare, valutare e generare informazioni in modo sicuro e adeguato tramite strumenti digitali e tecnologie in rete per prendere parte alla vita economica e sociale.

Algoritmo
Un algoritmo è un metodo di risoluzione dei problemi che utilizza una serie di operazioni finita e non ambigua. Specificamente, nel contesto dell'intelligenza artificiale, indica una serie di operazioni utilizzate per inserire dati e raggiungere il risultato desiderato.

Apprendimento approfondito
L'apprendimento approfondito è la branca dell'apprendimento automatico che utilizza reti di neuroni artificiali a vari livelli. Questa tecnologia è alla base delle più recenti innovazioni di IA.

Apprendimento automatico
L'apprendimento automatico è la branca dell'intelligenza artificiale che consiste nella programmazione di un algoritmo in modo che sia in grado di apprendere autonomamente.

Le diverse tecnologie possono essere classificate secondo tre tipi principali di apprendimento automatico:

> Nell'apprendimento supervisionato, i sistemi IA imparano a prevedere un valore dai dati inseriti. Questo processo richiede coppie voce-valore annotate durante il training. Ad esempio, un sistema può imparare a riconoscere un oggetto raffigurato in un'immagine.

> Nell'apprendimento supervisionato, i sistemi IA imparano a trovare somiglianze tra i dati non annotati, ad esempio per dividerli in varie porzioni omogenee. I sistemi sono quindi in grado di riconoscere le comunità di utenti dei social media.

> Grazie all'apprendimento per rinforzo, i sistemi IA imparano ad agire sull'ambiente per aumentare al massimo i benefici ricevuti dal training. Questa tecnica ha consentito al sistema IA di battere gli esseri umani nel gioco Go o nel videogioco Dota2.

Attività online
L'attività online indica tutte le attività eseguite da un individuo in un ambiente digitale, tramite computer, telefono o altri oggetti collegati.

Beni digitali
I beni digitali sono le applicazioni o i dati prodotti da una comunità. Contrariamente ai beni materiali, sono facilmente condivisibili e non si deteriorano con l'uso. Pertanto, contrariamente ai software proprietari, sono considerati beni digitali i software open source (spesso risultanti dalla collaborazione tra programmatori), poiché dispongono di codice sorgente aperto e accessibile a tutti.

Bolla di filtraggio
L'espressione bolla di filtraggio fa riferimento alle informazioni "filtrate" che un individuo riceve da Internet. Vari servizi come i social network o i motori di ricerca offrono ai propri utenti risultati personalizzati. Di conseguenza, le persone possono risultare isolate (all'interno di una "bolla"), in quanto non hanno più accesso alle informazioni comuni.

Chatbot
Un chatbot è un sistema IA in grado di conversare con l'utente in una lingua naturale.

Dati aperti
I dati aperti sono dati digitali a cui gli utenti possono accedere liberamente. Ad esempio, la maggior parte dei risultati di ricerca IA pubblicati.

Dati personali
I dati personali sono i dati che contribuiscono, direttamente o indirettamente, all'identificazione di una persona.

Disconnessione digitale
La disconnessione digitale indica la sospensione temporanea o permanente dell'attività online di un utente.

Dipendenza del percorso
Meccanismo sociale attraverso il quale le decisioni tecnologiche in precedenza ritenute razionali ma ora sotto standard, continuano a influenzare le decisioni. Si tratta di un meccanismo mantenuto a fronte di preconcetti cognitivi o perché un cambiamento richiederebbe troppi fondi o troppo sforzo. È questo il caso delle infrastrutture stradali urbane, quando porta a programmi per l'ottimizzazione del traffico anziché prendere in considerazione una diversa organizzazione dei trasporti con veicoli a emissioni di carbonio ridotte. Questo meccanismo deve essere noto durante l'utilizzo dei sistemi IA per progetti speciali, poiché i dati di training nell'apprendimento supervisionato possono a volte consolidare vecchi paradigmi organizzativi che vengono ora contestati.

Effetto rimbalzo
L'effetto rimbalzo indica il meccanismo per il quale una migliore efficienza energetica o migliori prestazioni ambientali di prodotti, apparecchiature e servizi incrementano l'utilizzo in modo più che proporzionale. Ad esempio, aumentano le dimensioni degli schermi, cresce il numero di dispositivi elettronici di una famiglia e aumentano le distanze percorse in auto o aereo.
L'effetto a livello globale consiste in una maggiore pressione sulle risorse e sull'ambiente.

Forte sostenibilità ambientale
Il concetto di forte sostenibilità ambientale si riconduce all'idea che, per essere sostenibile, il consumo di risorse naturali e delle emissioni inquinanti deve essere compatibile con i limiti ambientali del pianeta, la quantità di rinnovamento di ecosistemi e risorse e la stabilità climatica.

Contrariamente alla sostenibilità debole, che richiede uno sforzo minore, la sostenibilità forte non consente di compensare la perdita di risorse naturali con capitale artificiale.

GAN
Acronimo di Generative Adversarial Network (Rete generativa avversaria). Una GAN mette in competizione due reti antagoniste per generare un'immagine. Può ad esempio essere utilizzata per creare un'immagine, una registrazione o un video che appaia praticamente reale a un essere umano.

Giustificabilità decisionale
Le decisioni dei sistemi IA sono giustificate quando sussistono motivi non futili che le motivano e che possono essere comunicati in una lingua naturale.

Intelligenza artificiale (IA)
L'intelligenza artificiale (IA) indica una serie di tecnologie che consentono alle macchine di simulare l'apprendimento umano, nella fattispecie di apprendere, prevedere, decidere e percepire l'ambiente circostante. Nel caso di un sistema di calcolo, l'intelligenza artificiale viene applicata ai dati digitali.

Intelligibilità

I sistemi IA sono intelligibili quando un essere umano dotato delle competenze necessarie è in grado di comprenderne le operazioni, vale a dire il modello matematico e i processi che le determinano.

Sistema di acquisizione e archiviazione dei dati (SAAD).

SAAD indica tutti i sistemi di calcolo in grado di raccogliere e registrare dati. Successivamente, i dati vengono utilizzati per effettuare il training dei sistemi IA oppure come parametri decisionali.

Sistema di intelligenza artificiale (SIA)

Un SIA è un sistema di calcolo che utilizza gli algoritmi dell'intelligenza artificiale, che si tratti di software, di oggetti collegati o di robot.

Sviluppo sostenibile

Lo sviluppo sostenibile indica uno sviluppo della società umana compatibile con la capacità dei sistemi naturali di offrire le risorse e i servizi necessari alla società stessa. Si tratta di uno sviluppo economico-sociale che rispetta le esigenze attuali senza compromettere l'esistenza delle future generazioni.

Training

Il training è il processo di apprendimento meccanico mediante il quale i sistemi IA generano un modello dai dati. Le prestazioni dei sistemi IA dipendono dalla qualità di tale modello, mentre quest'ultimo dipende dalla qualità e quantità dei dati utilizzati per il training.

CREDITI

Il testo della Dichiarazione di Montréal per lo sviluppo responsabile dell'intelligenza artificiale è il risultato del lavoro di un team scientifico interuniversitario e multidisciplinare, basato su un processo di consultazione dei cittadini e sul dialogo con esperti e soggetti interessati allo sviluppo dei sistemi IA.

Christophe Abrassart, professore associato della School of Design e co-direttore del Lab Ville Prospective della Facoltà di programmazione dell'Université de Montréal, membro del *Centre de recherche en éthique* (CRÉ)

Yoshua Bengio, professore ordinario del Dipartimento di Informatica e ricerca operativa, UdeM, direttore scientifico di MILA e IVADO

Guillaume Chicoisne, direttore programmi scientifici, IVADO

Nathalie de Marcellis-Warin, professore ordinario, Polytechnique Montréal, presidente direttore generale, Center for Interuniversity Research and Analysis of Organizations (CIRANO)

Marc-Antoine Dilhac, professore associato, Dipartimento di Filosofia, Université de Montréal, Presidenza del Gruppo di etica e politica, *Centre de recherche en éthique* (CRÉ), Canada Presidenza di Ricerca in etica pubblica e teoria politica, direttore dell'Institut Philosophie Citoyenneté Jeunesse

Sébastien Gambs, professore di informatica dell'Université du Québec à Montréal, Presidenza di Ricerca canadese in analisi etica di salvaguardia della privacy di Big Data

Vincent Gautrais, professore ordinario, Facoltà di Legge, Université de Montréal; direttore del *Centre de recherche en droit public* (CRDP); Presidenza L.R. Wilson di Legge su e-commerce e informatica

Martin Gibert, consulente di etica presso IVADO e ricercatore presso il *Centre de recherche en éthique* (CRÉ)

Lyse Langlois, professore ordinario e vice-preside della Facoltà di Scienze sociali; Direttore dell'*Institut d'éthique appliquée* (IDÉA); ricercatore dell'Interuniversity Research Center on Globalization and Work (CRIMT)

François Laviolette, professore ordinario, Dipartimento di Informatica e Ingegneria di software, Université Laval; direttore del Centre de recherche en données massives (CRDM)

Pascale Lehoux, professore ordinario presso l'*École de santé publique*, Université de Montréal (ESPUM); Presidenza di Innovazione responsabile e salute

Jocelyn Maclure, professore ordinario, Facoltà di Filosofia, Université Laval, e presidente di Quebec Ethics in Science and Technology Commission (CEST)

Marie Martel, professore presso l'*École de bibliothéconomie et des sciences de l'information*, Université de Montréal

Joëlle Pineau, professore associato, School of Computer Science, McGill University; direttore del Facebook AI Lab in Montréal; co-direttore del Reasoning and Learning Lab

Peter Railton, insigne professore universitario Gregory S. Kavka; professore John Stephenson Perrin; professore Arthur F. Thurnau, Dipartimento di Filosofia, University of Michigan, membro dell'American Academy of Arts & Sciences

Catherine Régis, professore associato, Facoltà di Legge, Université de Montréal; Presidenza della ricerca canadese in Cultura collaborativa nella politica e legge sanitaria; ricercatore ordinario, *Centre de recherche en droit public* (CRDP)

Christine Tappolet, professore ordinario, Dipartimento di Filosofia, UdeM, direttore del Centre de recherche en éthique (CRÉ)

Nathalie Voarino, dottorato di ricerca in Bioetica dell'Université de Montréal

I NOSTRI PARTNER

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

Mila

ICRA
Programme
IA et
société

Université de Montréal | designNsociété

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

Québec
Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture

UNIVERSITÉ LAVAL
Institut d'éthique appliquée

MUSÉE DE LA CIVILISATION
Québec

CENTRE D'ÉTUDES ET DE RECHERCHES INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE EN DONNÉES MASSIVES DE L'UNIVERSITÉ LAVAL

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

Montréal Declaration Responsible AI_

montrealdeclaration-responsibleai.com