

La fratellanza dell'era della AI

Il nostro appello globale per una pacifica coesistenza umana e una condivisione delle responsabilità

a Sua Santità Papa Leone XIV

a tutti i leader globali

a tutte le persone di buona volontà

Animati dal profondo desiderio di un futuro in cui siano gli esseri umani a guidare la società e le sue scelte, noi, un gruppo di lavoro indipendente composto da esperti, leader tecnologici, intellettuali e studiosi provenienti da diverse nazioni, culture e fedi, lanciamo questo appello affinché l'intelligenza artificiale sia sviluppata responsabilmente, da e per le persone.

Le scelte che facciamo oggi nell'ambito dell'intelligenza artificiale incideranno radicalmente sul mondo che lasceremo alle future generazioni. L'IA sta già causando danni significativi: accentua le disuguaglianze, concentra il potere nelle mani di pochi attori e produce effetti nocivi sull'ambiente. Somme ingenti e sempre più cospicue sono destinate allo sviluppo di tecnologie agentiche con il potenziale di superare le capacità cognitive umane: ciò che molti esponenti della ricerca nell'ambito dell'IA definiscono "superintelligenza". Queste sfide richiedono una leadership morale e interventi concreti e tempestivi.

L'intelligenza artificiale offre opportunità significative per accelerare la ricerca scientifica, favorire una più profonda comprensione tra gli esseri umani, rivoluzionare l'assistenza sanitaria, rafforzare i modelli di governance e, più in generale, contribuire alla costruzione di una prosperità sostenibile e inclusiva. Tuttavia, l'intelligenza artificiale comporta anche rischi significativi, come evidenziato nell'International Scientific Report on AI Safety: dalla perdita di posti di lavoro alla riduzione delle libertà individuali, dalle guerre di potere alla disinformazione e manipolazione, fino alla sorveglianza di massa, agli impatti ambientali e a più ampie minacce al benessere umano.

Al fine di cogliere appieno le opportunità legittime e potenziali, mitigando al contempo costi e rischi, è fondamentale porre le basi per la prosperità umana, accompagnata da confini chiari e saldi, radicati nel rispetto della dignità, della comunità, dei diritti umani e ambientali e fondati sul principio di responsabilità.

In questo spirito di fraternità, speranza e cautela, rivolgiamo un appello alla vostra leadership affinché sostenga i seguenti principi e limiti invalicabili, con l'obiettivo di

favorire il dialogo e la riflessione su come l'intelligenza artificiale possa servire al meglio l'intera umanità.

Vita e dignità umane: L'intelligenza artificiale non deve mai essere sviluppata né impiegata in modi che possano minacciare, sminuire o negare la vita, la dignità o i diritti fondamentali della persona. Intelligenza umana: la nostra facoltà di giudizio, di ragionamento morale e l'orientamento verso la verità e la bellezza non devono mai essere svalutati da alcuna forma di elaborazione artificiale, per quanto sofisticata.

L'IA deve essere utilizzata come strumento, non come autorità: l'IA deve rimanere sotto il controllo umano. Costruire sistemi incontrollabili o delegare eccessivamente le decisioni è moralmente inaccettabile e deve essere vietato dalla legge. Pertanto, lo sviluppo di tecnologie di intelligenza artificiale di superintelligenza (come menzionato sopra) non deve essere autorizzato finché non sarà raggiunto un ampio consenso scientifico sul fatto che esse possano essere realizzate in modo sicuro e controllabile, e finché non sarà raggiunto un chiaro consenso pubblico.

Responsabilità: solo gli esseri umani possiedono lo status di soggetti morali e giuridici; i sistemi di intelligenza artificiale sono e devono rimanere oggetti di diritto, mai soggetti. Le responsabilità e gli obblighi ricadono su sviluppatori, fornitori, aziende, implementatori, utenti, istituti e governi. All'IA non può essere concessa la personalità giuridica o "diritti".

Decisioni di vita o di morte: ai sistemi di intelligenza artificiale non deve mai essere consentito di prendere decisioni di vita o di morte, specialmente nelle applicazioni militari durante i conflitti armati o in tempo di pace, nell'ambito dell'ordine pubblico, nel controllo delle frontiere, nell'assistenza sanitaria o nelle decisioni giudiziarie.

Sviluppo sicuro ed etico: gli sviluppatori devono progettare l'IA con sicurezza, trasparenza ed etica al centro, non come un elemento secondario. Gli utilizzatori devono considerare il contesto di utilizzo e i potenziali danni e sono soggetti agli stessi principi etici e di sicurezza degli sviluppatori. Prima della distribuzione e durante l'intero ciclo di vita devono essere richiesti test indipendenti e un'adeguata valutazione del rischio.

Custodia e responsabilità: i governi, le imprese e chiunque altro non dovrebbero trasformare l'IA in uno strumento di dominio, di guerra di aggressione illegale, coercizione, manipolazione, social scoring o sorveglianza di massa ingiustificata.

Progettazione responsabile: l'IA dovrebbe essere progettata e valutata in modo indipendente per evitare effetti imprevisti e catastrofici sugli esseri umani e sulla società, ad esempio attraverso una progettazione che possa generare inganno, illusione, dipendenza o perdita di autonomia.

Nessun monopolio dell'IA: i benefici dell'IA (economici, medici, scientifici, sociali)

non devono essere monopolizzati.

Nessuno svilimento umano: la progettazione e l'implementazione dell'IA dovrebbero far prosperare gli esseri umani nelle loro attività elettive, non rendere l'umanità ridondante, estromessa, svalutata o sostituibile.

Responsabilità ecologica: il nostro uso dell'IA non deve mettere in pericolo il nostro pianeta e gli ecosistemi. Le vaste richieste di energia, acqua e terre rare dell'IA devono essere gestite in modo responsabile e sostenibile lungo tutta la catena di approvvigionamento.

Nessuna concorrenza globale irresponsabile: dobbiamo evitare una corsa irresponsabile tra aziende e paesi verso un'IA sempre più potente.

Rispettare questi principi non sarà facile. Richiede coraggio morale, una responsabilizzazione efficace, una leadership lungimirante da parte di tutti i settori della società e un trattato internazionale vincolante che stabilisca i limiti invalicabili e un'istituzione di controllo indipendente con poteri coercitivi e sanzionatori. Pertanto, chiediamo una leadership morale per l'era dell'IA. Poiché i pericoli presentati dall'IA sono spesso indiretti, chiediamo agli scienziati, alla società civile, ai gruppi per i diritti e ad altre parti interessate di compiere uno sforzo maggiore per articolare, e amplificare la consapevolezza pubblica, i limiti e i pericoli dell'IA. Invitiamo scienziati, leader del settore tecnologico e responsabili politici ad ascoltare le voci, le esperienze e la ricerca di chi lavora con sui dati, delle comunità e delle persone che subiscono le conseguenze (del lato materiale) dell'IA. Gli invitiamo a concentrare il loro lavoro sulla protezione e il beneficio dei più vulnerabili. Perché la legittimità delle norme morali e legali in una società dipende da come tratta le persone più vulnerabili.

Facciamo inoltre appello a scienziati, alla società civile e revisori indipendenti affinché sviluppino e proponano nuovi obiettivi e metriche per addestrare, ottimizzare e valutare gli algoritmi di apprendimento in termini di veridicità, equilibrio e bene umano, durante l'intero ciclo di vita, e non solo in termini di prestazioni e engagement.

Incoraggiamo i legislatori, i leader del settore tecnologico e le comunità globali a collaborare per sviluppare linee guida esaurienti per la governance dell'IA che servano il bene comune. Ciò include il diritto degli esseri umani a vivere senza IA. Il progresso di una vera fraternità umana nell'era dell'intelligenza artificiale richiede l'istituzione di standard etici e legali universali.

Infine, ci appelliamo a tutte le persone di buona volontà: uniamoci per garantire che l'IA serva tutta l'umanità piuttosto che una ristretta minoranza.

Unendoci attraverso nazioni, culture e credi, dando priorità al dialogo rispetto alla

concorrenza, possiamo plasmare un futuro che elevi la dignità umana e promuova un mondo più giusto e pacifico.

Invitiamo tutte le parti interessate, inclusi cittadini, scienziati, leader aziendali, leader religiosi, rappresentanti della comunità e responsabili politici, a partecipare a questa iniziativa. Insieme, ribadiamo il principio essenziale secondo cui le macchine devono servire gli interessi dell'umanità.