VareseNews

Intelligenza artificiale, Harari: “Sappiamo di non sapere“

Pubblicato: Sabato 11 Novembre 2023

La settimana scorsa ero a Londra con la famiglia. Mentre stavamo visitando l’area della residenza reale, ci siamo trovati faccia a faccia con Louis, il più giovane figlio di William e Kate, quarto in linea di successione, che stava uscendo da Buckingham Palace con l’autista (foto in basso). Una fugace quanto inattesa fortuna da turisti, inconsapevoli che in quel momento c’era qualcosa di molto più significativo che stava accadendo nella capitale del Regno Unito: **la firma della dichiarazione di Bletchley**.

Questa dichiarazione è la prima sottoscritta a livello internazionale dai 28 Paesi, che hanno partecipato al summit sull’Intelligenza Artificiale (IA), tenutosi a Londra l’ 1 e 2 novembre 2023. In breve, la dichiarazione sottolinea le vaste opportunità offerte dall’IA per la trasformazione globale, il benessere e la prosperità.

Pur riconoscendone l’impatto positivo, **la dichiarazione mette altresì in evidenza i significativi rischi che essa comporta.** Quindi, sottolinea la necessità di una cooperazione internazionale per affrontare questi rischi e promuovere l’uso responsabile dell’IA.

Si presta particolare attenzione alla sicurezza delle capacità delle tecnologie IA di frontiera, quelle più avanzate, con un’enfasi sulla trasparenza, responsabilità e politiche basate sulla prevenzione, controllo e gestione dei rischi. La dichiarazione è un risultato non scontato, soprattutto perché **sottoscritta da Paesi che non sono particolarmente allineati nel contesto geopolitico** globale come USA, Cina, India, Arabia Saudita e Turchia, oltre che dall’Italia. Vediamo in dettaglio cosa viene sottolineato.

Principi fondamentali.

“L’Intelligenza Artificiale (IA) presenta enormi opportunità globali: ha il potenziale per trasformare e migliorare il benessere umano, la pace e la prosperità. Per realizzare ciò, affermiamo che, a beneficio di tutti, l’IA dovrebbe essere progettata, sviluppata, implementata e utilizzata in modo sicuro e, affinché metta al centro l’umano, in modalità affidabili e responsabili”.

Rischi.

“Particolari rischi di sicurezza emergono alla ‘frontiera’ dell’IA … Rischi significativi possono derivare sia da un possibile uso intenzionale improprio o da problemi non intenzionali che emergono dalla perdita di controllo dall’intento di fare del bene all’uomo. Questi problemi sono in parte dovuti al fatto che tali capacità non sono completamente comprese e quindi difficili da prevedere. Siamo particolarmente preoccupati per tali rischi in settori come la cibersicurezza e la biotecnologia, nonché dove i sistemi di IA di frontiera possono amplificare rischi come la disinformazione. C’è il potenziale per danni gravi, addirittura catastrofici, sia intenzionali che non intenzionali, derivanti dalle capacità più significative di questi modelli di IA. Dato il rapido e incerto tasso di cambiamento dell’IA e nel contesto dell’accelerazione degli investimenti nella tecnologia, affermiamo che approfondire la comprensione di questi rischi potenziali e delle azioni per affrontarli è particolarmente urgente”.

Risoluzioni.

“Nel contesto della nostra cooperazione e per dare forma all’azione a livello nazionale e internazionale, la nostra agenda per affrontare i rischi dell’IA di frontiera si concentrerà su:
Identificare i rischi di sicurezza legati all’IA di comune preoccupazione, costruire una comprensione condivisa basata su evidenze scientifiche di questi rischi e mantenere tale comprensione man mano che le capacità continuano a crescere, nel contesto di un approccio globale più ampio per comprendere l’impatto dell’IA nelle nostre società.
Sviluppare politiche basate sul rischio nei nostri rispettivi paesi per garantire la sicurezza alla luce di tali rischi, collaborando come opportuno e riconoscendo che i nostri approcci possono differire in base alle

circostanze nazionali e ai quadri giuridici applicabili. Ciò include, insieme a una maggiore trasparenza da parte degli attori privati che sviluppano capacità di IA di frontiera, adeguati indicatori di valutazione, strumenti per i test di sicurezza e lo sviluppo di capacità rilevanti nel settore pubblico e nella ricerca scientifica.
Per promuovere questa agenda, ci impegniamo a sostenere una rete scientifica internazionalmente inclusiva sulla sicurezza dell'IA di frontiera che comprenda e completi la collaborazione esistente e nuova a livello multilaterale, plurilaterale e bilaterale, incluso attraverso i già esistenti fori internazionali e altre iniziative rilevanti, per facilitare la fornitura della migliore scienza disponibile per la formulazione di politiche e il bene pubblico"

Intervistato a margine della conferenza, cui ha partecipato**, Yuval Noah Harari**, lo storico, autore di "Sapiens. Da animali a dèi. Breve storia dell'umanità", ha affermato che l'IA potrebbe causare una crisi finanziaria con conseguenze "catastrofiche" a causa della difficoltà nel prevederne i pericoli. "L'IA è diversa da ogni tecnologia precedente nella storia umana perché è la prima tecnologia capace di prendere decisioni autonomamente, di generare nuove idee in modo autonomo e di imparare e svilupparsi autonomamente. Quasi per definizione, è estremamente difficile per gli esseri umani, persino per coloro che hanno creato la tecnologia, prevedere tutti i potenziali pericoli e problemi".

A differenza delle armi nucleari, l'IA presenta numerosi scenari pericolosi, ciascuno con una piccola probabilità, che insieme costituiscono una minaccia per la civiltà umana. Harari elogia la cooperazione globale espressa nell'ultimo vertice sulla sicurezza dell'IA, ma sottolinea la sfida unica rappresentata dalla capacità decisionale autonoma dell'IA e la necessità di istituzioni regolatorie potenti e tempestive per affrontare pericoli imprevedibili. Lo storico israeliano evidenzia il settore finanziario come particolarmente suscettibile a crisi generate dall'IA, immaginando scenari in cui l'IA crea dispositivi finanziari complessi al di là della comprensione umana, riflettendo le preoccupazioni sollevate da vari governi. Harari sottolinea l'importanza di istituzioni regolatorie con competenze in materia di IA, sostenendo un approccio proattivo piuttosto che fare affidamento su regolamentazioni lunghe e obsolete.

La narrazione dominante che viene rafforzata da questa notizia può essere riassunta così: "L'IA, con le sue immense opportunità, porta con sé rischi significativi: cooperazione internazionale e regolamentazioni tempestive sono essenziali per garantire un futuro sicuro e responsabile". Bene. Ma forse è giunto il momento di elevarci oltre la mera gestione tecnologica e approfondire la riflessione su quale vita, valori e concezione dell'umanità desideriamo preservare e sviluppare. La vera sfida di questa, e della tecnologia in generale, è filosofica e culturale.
"È sapiente solo chi sa di non sapere, non chi s'illude di sapere e ignora così perfino la sua stessa ignoranza", Socrate.

di Giuseppe Geneletti