

Intelligenza artificiale spiegabile e solida (Partenariato per i dati e la robotica dell'intelligenza artificiale) **Explainable and Robust AI (AI Data and Robotics Partnership)**

TOPIC ID:

HORIZON-CL4-2024-HUMAN-01-06

Ente finanziatore:

Commissione europea - Programma Horizon Europe

Obiettivi ed impatto attesi:

I progetti devono contribuire a uno dei seguenti risultati:

- Miglioramento della robustezza, delle prestazioni e dell'affidabilità dei sistemi di IA, compresa la consapevolezza dei limiti della robustezza operativa del sistema.
- Miglioramento della spiegabilità e della responsabilità, della trasparenza e dell'autonomia dei sistemi di IA, compresa la consapevolezza delle condizioni di lavoro del sistema.

Ambito di applicazione:

Le soluzioni di IA degne di fiducia devono essere robuste, sicure e affidabili quando operano in condizioni reali e devono essere in grado di fornire spiegazioni adeguate, significative e complete, se pertinenti, o approfondimenti sulla causalità, tenere conto delle preoccupazioni relative all'equità, essere solide quando si affrontano tali questioni in condizioni reali, rispettando al contempo i diritti e gli obblighi relativi all'uso dei sistemi di IA in Europa. I progressi in queste aree possono contribuire a creare un'IA incentrata sull'uomo[1] che rifletta le esigenze e i valori dei cittadini europei e contribuisca a una governance efficace delle tecnologie di IA.

Per ottenere un'IA robusta e affidabile, sono necessari nuovi approcci per sviluppare metodi e soluzioni che funzionino in circostanze diverse da quelle ideali del modello e che siano consapevoli quando queste condizioni vengono meno. Per ottenere l'affidabilità, i sistemi di IA devono essere sufficientemente trasparenti e in grado di spiegare come il sistema ha raggiunto una conclusione in modo che sia significativa per l'utente, indicando anche quando sono stati raggiunti i limiti di funzionamento.

Lo scopo è quello di far progredire algoritmi di intelligenza artificiale in grado di operare in modo sicuro in una varietà comune di circostanze, in modo affidabile in condizioni reali e di prevedere quando queste circostanze operative non sono più valide. La ricerca deve mirare a far progredire la robustezza e la spiegabilità per una generalità di soluzioni, pur comportando una perdita accettabile in termini di accuratezza ed efficienza, e con una verificabilità e riproducibilità note. L'obiettivo è estendere l'applicabilità generale della spiegabilità e della robustezza dei sistemi di IA attraverso la ricerca fondamentale sull'IA e sull'apprendimento automatico. A tal fine, possono essere presi in considerazione i seguenti metodi, ma non sono necessariamente limitati a:

- apprendimento efficiente dei dati, trasformatori, apprendimento per rinforzo, apprendimento federato ed edge-learning, apprendimento automatico o qualsiasi combinazione di questi elementi per migliorare la robustezza e la spiegabilità.

- approcci ibridi che integrano apprendimento, conoscenza e ragionamento, approcci basati su modelli, calcolo neuromorfico o altri approcci ispirati alla natura e altre forme di combinazioni ibride genericamente applicabili alla robustezza e alla spiegabilità.
- apprendimento continuo, apprendimento attivo, apprendimento a lungo termine e come possono contribuire a migliorare la robustezza e la spiegabilità.
- l'apprendimento multimodale, l'elaborazione del linguaggio naturale, il riconoscimento vocale e la comprensione del testo tenendo conto degli aspetti multiculturali allo scopo di aumentare la robustezza operativa e la capacità di spiegare formulazioni alternative.[2].

Le attività di ricerca multidisciplinari devono riguardare tutti i seguenti aspetti:

- Le proposte devono coinvolgere competenze adeguate in tutte le discipline pertinenti e, se del caso, nelle scienze sociali e umanistiche (SSH), comprese le conoscenze di genere e intersezionali per affrontare le preoccupazioni relative a pregiudizi di genere, razziali o di altro tipo, ecc.
- Le proposte devono dedicare compiti e risorse per collaborare e fornire input alla sfida di innovazione aperta nell'ambito di HORIZON-CL4-2023-HUMAN-01-04 che riguarda la spiegabilità e la robustezza. I gruppi di ricerca coinvolti nelle proposte devono partecipare alle rispettive sfide dell'innovazione.
- Contribuire a far sì che le soluzioni di IA e robotica soddisfino i requisiti di un'IA affidabile, basata sul rispetto dei principi etici, dei diritti fondamentali e di aspetti critici quali robustezza, sicurezza e affidabilità, in linea con l'approccio europeo all'IA. I principi etici devono essere adottati fin dalle prime fasi di sviluppo e progettazione.

Tutte le proposte devono incorporare meccanismi per valutare e dimostrare i progressi (con KPI qualitativi e quantitativi, benchmarking e monitoraggio dei progressi) e condividere i risultati comunicabili con la comunità europea di R&S, attraverso la piattaforma AI-on-demand o la piattaforma industriale digitale per la robotica, le risorse pubbliche della comunità, per massimizzare il riutilizzo dei risultati, sia da parte degli sviluppatori, sia per l'adozione, e ottimizzare l'efficienza dei finanziamenti; migliorare l'ecosistema europeo dell'AI, dei dati e della robotica attraverso la condivisione dei risultati e delle migliori pratiche.

Per raggiungere i risultati previsti, si incoraggia la cooperazione internazionale, in particolare con il Canada e l'India.

Condizioni specifiche dell'argomento:

Le attività dovrebbero iniziare a TRL 2-3 e raggiungere TRL 4-5 entro la fine del progetto.

Criteri di eleggibilità:

Condizioni di ammissibilità

Le condizioni sono descritte nell'Allegato generale B.

Per essere ammissibili, i richiedenti (beneficiari ed enti affiliati) devono:

-essere soggetti giuridici (enti pubblici o privati) avere sede in uno dei Paesi ammissibili, ovvero Stati membri dell'UE (compresi i Paesi e territori d'oltremare (PTOM)) Paesi non UE:

- Paesi SEE elencati e Paesi associati o Paesi che hanno in corso negoziati per un accordo di associazione e

in cui l'accordo entra in vigore prima della firma della sovvenzione (elenco dei Paesi partecipanti)

– Paesi in via di adesione,

I beneficiari e gli enti affiliati devono iscriversi al Registro dei partecipanti – prima di presentare la proposta – e dovranno essere convalidati dal Servizio centrale di convalida (REA Validation). Per la convalida, sarà loro richiesto di caricare documenti che dimostrino lo status giuridico e l'origine.

Si applicano le seguenti eccezioni: Se i progetti utilizzano dati e servizi di osservazione della terra, posizionamento, navigazione e/o tempistica correlati basati su satelliti, i beneficiari devono utilizzare Copernicus e/o Galileo/EGNOS (possono essere utilizzati anche altri dati e servizi).

Contributo finanziario:

Contributo UE previsto per progetto La Commissione ritiene che un contributo UE compreso tra 9 e 10 milioni di euro consentirebbe di affrontare adeguatamente questi risultati. Tuttavia, ciò non preclude la presentazione e la selezione di una proposta che richieda importi diversi.

Il budget totale indicativo per il tema è di 30,00 milioni di euro. Tipo di azione : Azioni di ricerca e innovazione

Scadenza:

19 marzo 2024 17:00:00 ora di Bruxelles

Ulteriori informazioni:

[wp-7-digital-industry-and-space_horizon-2023-2024_en.pdf \(europa.eu\)](#)