Intelligenza Artificiale comprensibile e solida (Partenariato per i dati e la robotica dell'intelligenza artificiale) (RIA)

Explainable and Robust AI (AI Data and Robotics Partnership) (RIA)

TOPIC ID:
HORIZON-CL4-2024-HUMAN-03-02

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, compresi i modelli di IA generativa, con la consapevolezza dei limiti della robustezza operativa del sistema.
- Miglioramento della spiegabilità e della responsabilità, della trasparenza e dell'autonomia dei sistemi di IA, compresi i modelli di IA generativa, insieme alla 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.
La necessità di sistemi di IA trasparenti e robusti è diventata più pressante con la rapida crescita e commercializzazione di sistemi di IA generativi basati su modelli di fondazione. Nonostante le loro impressionanti capacità, l'affidabilità rimane una sfida scientifica fondamentale e irrisolta. A causa della natura intricata dei sistemi di IA generativa, la comprensione o la spiegazione della logica alla base dei loro risultati non è normalmente possibile con gli attuali metodi di IA spiegabili. Inoltre, questi modelli tendono occasionalmente ad avere "allucinazioni", generando informazioni non reali o inaccurate, compromettendo ulteriormente la loro affidabilità.
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, pur essendo 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 è giunto a una conclusione in modo significativo per l'utente, consentendo un'interazione uomo-macchina sicura e protetta e indicando al contempo quando sono stati raggiunti i limiti di funzionamento.
L'obiettivo è far progredire gli algoritmi di intelligenza artificiale e le innovazioni basate su di essi, in grado di funzionare in modo sicuro in una varietà comune di circostanze, in modo affidabile nelle condizioni del mondo reale e di prevedere quando queste circostanze operative non sono più valide. La ricerca dovrebbe mirare a migliorare 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 fondazionale 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 e architetture alternative, apprendimento auto-supervisionato, messa a punto dei modelli di base, 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 l'apprendimento, la conoscenza e il ragionamento, i metodi neurosimbolici, gli approcci basati su modelli, l'informatica neuromorfa 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 tutti i casi d'uso e le discipline specifiche del settore e, se del caso, nelle scienze sociali e umane (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 team 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; potenziare l'ecosistema europeo dell'AI, dei dati e della robotica ed eventuali forum settoriali specifici 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:
Si prevede che le attività inizino a TRL 2-3 e raggiungano TRL 4-5 entro la fine del progetto - cfr. Allegato generale B.

Criteri di eleggibilità:
Qualsiasi soggetto giuridico, indipendentemente dal suo luogo di stabilimento, compresi i soggetti giuridici di Paesi terzi non associati o di organizzazioni internazionali (comprese le organizzazioni internazionali di

ricerca europee) è ammesso a partecipare (indipendentemente dal fatto che sia ammissibile o meno al finanziamento), a condizione che siano state soddisfatte le condizioni stabilite dal regolamento Horizon Europe e qualsiasi altra condizione stabilita nel tema specifico del bando. Per “soggetto giuridico” si intende qualsiasi persona fisica o giuridica costituita e riconosciuta come tale ai sensi del diritto nazionale, del diritto dell’UE o del diritto internazionale, dotata di personalità giuridica e che può, agendo in nome proprio, esercitare diritti ed essere soggetta a obblighi, oppure un soggetto privo di personalità giuridica. I beneficiari e gli enti affiliati devono registrarsi nel Registro dei Partecipanti prima di presentare la domanda, per ottenere un codice di identificazione dei partecipanti (PIC) ed essere convalidati dal Servizio Centrale di Convalida prima di firmare la convenzione di sovvenzione. Per la convalida, durante la fase di preparazione della sovvenzione, verrà chiesto loro di caricare i documenti necessari che dimostrino il loro status giuridico e la loro origine. Un PIC convalidato non è un prerequisito per presentare una domanda.

Contributo finanziario:
Contributo UE previsto per progetto La Commissione stima che un contributo UE di circa 7,50 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. Budget indicativo Il budget totale indicativo per il tema è di 15,00 milioni di euro. Tipo di azione Azioni di ricerca e innovazione

Scadenza:
18 settembre 2024 17:00:00 ora di Bruxelles

Ulteriori informazioni:
wp-7-digital-industry-and-space_horizon-2023-2024_en.pdf (europa.eu)