

Quattro motivi per utilizzare i modelli linguistici di piccole dimensioni open source

I modelli linguistici di piccole dimensioni stanno trasformando le strategie aziendali legate all'IA

I modelli linguistici di grandi dimensioni (LLM) proprietari sono perfetti per le applicazioni di utilizzo generico, ma non sono sempre la scelta più adatta per le soluzioni di intelligenza artificiale (IA) aziendali. Sono caratterizzati infatti da requisiti di elaborazione piuttosto elevati, processi decisionali poco trasparenti e licenze costose, che possono limitare la flessibilità e aumentare la complessità operativa. I modelli linguistici di piccole dimensioni (SLM), in particolare se open source, offrono un'alternativa alle aziende che desiderano sviluppare soluzioni di IA personalizzate, mantenere il controllo sui dati e gestire i costi in modo efficiente.

Ecco quattro motivi per cui gli SLM open source potrebbero essere la scelta giusta per i tuoi progetti di IA.

1 Accesso all'innovazione della community

Combinando flessibilità, collaborazione e innovazione, gli SLM open source costituiscono una base per la creazione di applicazioni di IA altamente adattabili e specializzate. Poiché offrono l'accesso sia ai componenti software sia ai pesi del modello pre-addestrati, i progetti di IA open source consentono di collaborare con una community globale di sviluppatori e ricercatori per ottimizzare costantemente le tecnologie di IA generativa (IA gen). La condivisione dell'innovazione offre l'opportunità di usufruire di strumenti moderni e avanzati, personalizzandoli in base alle esigenze tecniche delle soluzioni di IA della propria azienda.

Nel caso degli SLM open source come i modelli di IA gen della [famiglia Granite di IBM](#), puoi contribuire in modo diretto allo sviluppo dei modelli fondativi con le tue conoscenze e competenze settoriali. Aniché aspettare gli aggiornamenti degli LLM proprietari, puoi personalizzare attivamente gli SLM open source per allinearli alle tue applicazioni di IA e migliorarne le prestazioni. Grazie a questo approccio interattivo puoi velocizzare l'iterazione e aggiornare i modelli seguendo l'evoluzione delle tue esigenze aziendali.

Gli SLM open source offrono la flessibilità necessaria per la distribuzione in ambienti dinamici, nei datacenter in loco e nell'infrastruttura di cloud pubblico. Mantenendo il pieno controllo dei tuoi modelli, puoi ottimizzarli per una serie di casi di deployment, dagli ambienti ad elevata conformità all'elaborazione dell'IA in tempo reale. E poiché contribuiscono a mantenere il controllo dello stack delle tecnologie di IA, gli SLM open source garantiscono che le tue soluzioni di IA innovative restino adattabili e scalabili man mano che le esigenze tecniche e aziendali si evolvono.

2 Controllo dei dati di addestramento

Gli SLM open source offrono maggiore trasparenza rispetto alle alternative proprietarie. I dati utilizzati per preaddestrare i modelli vengono divulgati da fornitori di fiducia, quindi puoi verificare in dettaglio la qualità dei modelli e assicurarti che non includano bias o informazioni dannose. Grazie a questa trasparenza puoi prendere decisioni informate sull'adattamento e sulla distribuzione dei modelli e, di conseguenza, assicurarti che le tue soluzioni di IA siano conformi agli standard etici e agli obiettivi aziendali, prima di integrare le informazioni sensibili di tua proprietà.

Inoltre, la possibilità di distribuire gli SLM in datacenter in loco e cloud privati all'interno degli ambienti IT aziendali consente di mantenere il pieno controllo sui dati di addestramento. Ciò è cruciale per le aziende che gestiscono dati altamente confidenziali o regolamentati, perché possono assicurarsi che le informazioni proprietarie non vengano mai esposte a provider esterni. In più, gestendo i modelli di IA gen all'interno del tuo ambiente, puoi controllare gli accessi, semplificare la conformità normativa, ottimizzare la sicurezza dei dati e aumentare la trasparenza nelle soluzioni di IA.

Infine, i modelli della famiglia Granite di IBM offrono [polizze di assicurazione](#) che proteggono i clienti in caso di reclami relativi a violazioni della proprietà intellettuale di terze parti da parte dei modelli di IA o dei software open source forniti. Scegliere questi modelli e fornitori può proteggere ulteriormente la tua azienda in un panorama tecnologico complesso e in continua evoluzione.

3 Personalizzazione delle soluzioni di IA

Con gli SLM open source puoi sviluppare soluzioni di IA in modo rapido ed efficiente, adattandole ai requisiti specifici della tua attività. Pensati e realizzati per scenari di utilizzo specifici, questi modelli offrono una soluzione mirata alle sfide tipiche del settore, senza la complessità e le ingenti risorse richieste dagli LLM a uso generico.

Ottimizzando gli SLM con i dati della tua organizzazione, puoi integrare le conoscenze aziendali e le competenze di settore direttamente nei parametri del modello. Questo approccio rende più pertinenti le risposte degli SLM, limita la frequenza e i costi del riaddestramento e riduce le tempistiche dello sviluppo di servizi e applicazioni di IA critici.

Rispetto agli LLM, gli SLM sono caratterizzati da dimensioni ridotte e richiedono volumi di dati minori, quindi sono più semplici da personalizzare e permettono di sviluppare modelli precisi ed efficienti che siano ottimizzati per attività o domini specifici. Inoltre, negli ambienti con risorse limitate e nei deployment all'edge, gli SLM consentono l'esecuzione di applicazioni in tempo reale direttamente sui dispositivi degli utenti, semplificando lo sviluppo ed eliminando la necessità di un'infrastruttura cloud esterna.

Gli SLM come i modelli Granite di IBM semplificano anche la transizione dalla fase di sperimentazione a quella di produzione. L'integrazione semplificata degli SLM con infrastrutture hardware e software di diverso tipo offre la possibilità di adattare le soluzioni di IA gen all'ambiente IT dell'azienda. Questo contribuisce a ridurre la complessità operativa, mantenendo al tempo stesso il controllo di distribuzioni e prestazioni.

4 Riduzione dei costi dei modelli di IA

Per molte aziende, ridurre i requisiti di elaborazione dell'IA è fondamentale per gestire in modo efficiente le spese. Gli SLM open source offrono le prestazioni necessarie per le soluzioni di IA gen avanzata, riducendo al tempo stesso sia i costi di addestramento e inferenza, sia la potenza di calcolo richiesta in confronto agli LLM.

Con dimensioni inferiori anche di migliaia di volte rispetto agli LLM più noti, gli SLM richiedono un volume decisamente ridotto di risorse, dati ed energia. Questa efficienza riduce i tempi di addestramento, semplifica il fine tuning e consente un approccio più sostenibile allo sviluppo dell'IA.

Inoltre, gli SLM open source sono scalabili in più progetti e organizzazioni senza richiedere dispendiosi upgrade dell'hardware. Integrando questi modelli nell'infrastruttura IT esistente, puoi creare soluzioni di IA personalizzate senza compromettere le prestazioni o superare i limiti di budget.

I risparmi sui costi non sono limitati all'infrastruttura. Gli SLM open source eliminano anche i costi delle licenze tipici dei modelli proprietari, offrendo quindi un accesso conveniente alle funzionalità di IA gen avanzate senza le limitazioni imposte dai fornitori.

Innova con gli SLM open source di Red Hat e IBM

I modelli di IA gen open source della famiglia Granite, sviluppati da IBM e inclusi in [Red Hat® Enterprise Linux® AI](#), rispondono alle esigenze specifiche delle applicazioni di IA aziendali.

Scopri di più sui modelli di IA gen open source

Leggi l'ebook [Ottimizza l'innovazione IA con i modelli open source](#) per sapere di più sugli SLM pensati per operazioni specifiche e sulle soluzioni di IA gen open source.



Informazioni su Red Hat

Red Hat consente la standardizzazione in diversi ambienti e lo sviluppo di applicazioni cloud native, oltre a favorire l'automazione, la protezione e la gestione di ambienti complessi grazie a [pluripremiati](#) servizi di consulenza, formazione e supporto.

f facebook.com/RedHatItaly
X twitter.com/RedHatItaly
in linkedin.com/company/red-hat

ITALIA
it.redhat.com
italy@redhat.com

**EUROPA, MEDIO ORIENTE,
E AFRICA (EMEA)**
00800 7334 2835
it.redhat.com
europe@redhat.com