

Piattaforma per modelli fondativi per l'IA generativa

Vantaggi principali

- ▶ Consente agli utenti di aggiornare e migliorare i modelli linguistici di grandi dimensioni (LLM) con InstructLab
- ▶ Allinea i LLM ai dati proprietari in tutta sicurezza per personalizzare i modelli in base alle specifiche esigenze aziendali
- ▶ Permette di iniziare a usare rapidamente l'intelligenza artificiale generativa (IA gen) e ottenete risultati con una piattaforma Red Hat® Enterprise Linux® affidabile e incentrata sulla sicurezza
- ▶ Sotto forma di immagine del container Red Hat Enterprise Linux di avvio per l'installazione e gli aggiornamenti

Panoramica sul prodotto

Red Hat Enterprise Linux AI è una piattaforma per modelli fondativi per l'intelligenza artificiale generativa di livello enterprise che consente di sviluppare, testare e distribuire LLM per gli scenari di utilizzo aziendali di IA gen.

Red Hat Enterprise Linux AI coniuga:

- ▶ La famiglia Granite di LLM open source.
- ▶ Strumenti di allineamento dei modelli InstructLab, che forniscono un approccio basato sulla community al fine tuning degli LLM.
- ▶ Un'immagine di avvio di Red Hat Enterprise Linux, insieme a librerie e dipendenze di IA gen, come PyTorch e software per driver acceleratori di IA per NVIDIA, Intel e AMD.
- ▶ Supporto tecnico a livello aziendale e indennizzi per tutelare la proprietà intellettuale dei modelli forniti da Red Hat.
- ▶ Red Hat Enterprise Linux AI offre la piattaforma affidabile Red Hat Enterprise Linux e aggiunge i componenti necessari per iniziare il percorso di utilizzo dell'IA generativa e vederne i risultati.

Il futuro dell'IA è aperto e trasparente

Red Hat Enterprise Linux AI comprende un sottoinsieme del linguaggio open source Granite e modelli di codice completamente tutelati da Red Hat. I modelli open source Granite forniscono alle organizzazioni modelli ottimizzati per costi e prestazioni che si allineano con un'ampia gamma di scenari di utilizzo dell'IA gen. I [modelli Granite](#) sono stati rilasciati su licenza Apache 2.0. Inoltre, non solo i modelli sono open source, ma anche i set di dati utilizzati per l'addestramento dei modelli sono trasparenti e aperti.

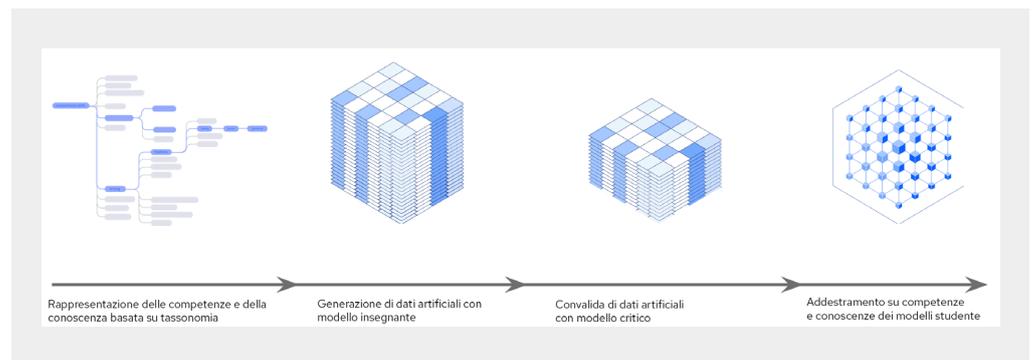
Tabella 1: modelli Granite in Red Hat Enterprise Linux AI

| | |
|------------------------------------|--|
| Modelli linguistici di IBM Granite | Granite-7B-Starter Granite-7B-RedHat-Lab |
| Modelli di codice di IBM | Granite-8B-Code-Instruct Granite-8B-Code-Base |

Addestramento dei modelli di IA gen accessibile per un time to value più rapido

Oltre ai modelli open source Granite, Red Hat Enterprise Linux AI include anche strumenti di allineamento dei modelli InstructLab, basati sulla tecnica di [allineamento su larga scala dei chatbot \(LAB\)](#). InstructLab consente ai team all'interno delle organizzazioni di contribuire in modo efficiente alle competenze e alle conoscenze dei LLM, personalizzando questi modelli in base alle esigenze specifiche della loro attività.

- ▶ Competenza: un dominio di capacità per insegnare a un modello come fare qualcosa. Le competenze sono classificate in due categorie:
 - ▶ Competenze compositive:
 - ▶ Permettono ai modelli di IA di svolgere attività o funzioni specifiche.
 - ▶ I modelli sono fondati (includono il contesto) o non fondati (non includono il contesto).
 - ▶ Un esempio di modello fondato è l'aggiunta di una competenza per fornire la capacità di leggere una tabella con formattazione Markdown.
 - ▶ Un esempio di modello non fondato è l'aggiunta di una competenza per insegnare al modello a comporre rime.
 - ▶ Competenze di base:
 - ▶ Abilità come la matematica, il ragionamento e la scrittura di codice.
- ▶ Conoscenze: dati e fact che forniscono a un modello dati e informazioni aggiuntivi per rispondere alle domande con maggiore precisione.



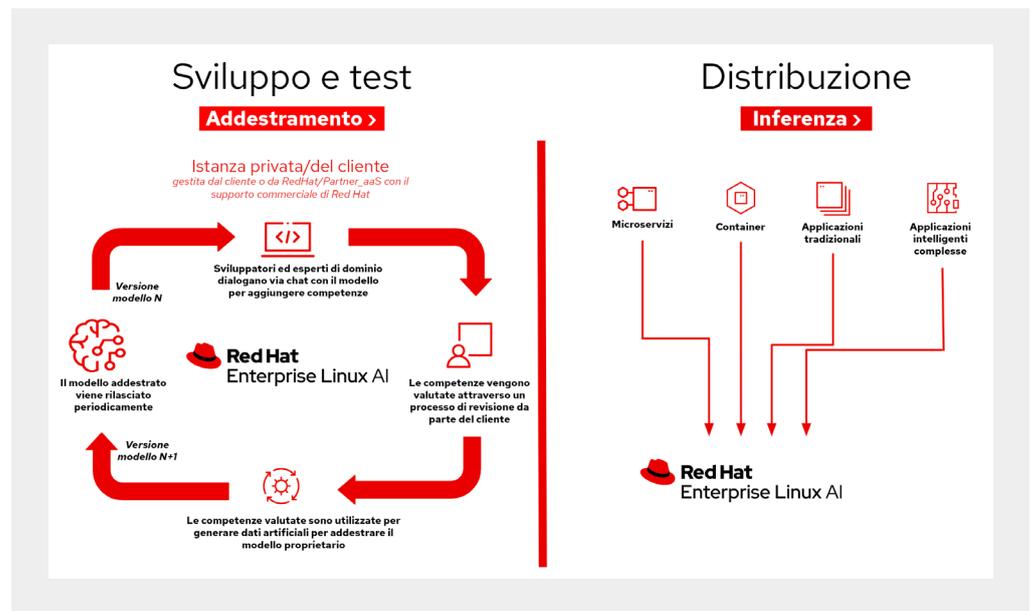
L'immagine qui sopra illustra il flusso di lavoro di fine tuning dei modelli InstructLab.

1. I contributi di competenze e conoscenze sono inseriti in un repository di dati basato su tassonomia.
2. Viene generata una quantità significativa di dati artificiali, utilizzando i dati della tassonomia, per produrre un set di dati sufficientemente grande da aggiornare e modificare correttamente un LLM.
3. L'output di dati artificiali viene rivisto, convalidato e sfoltito da un modello critico.
4. Il modello è addestrato con dati artificiali che hanno origine negli input manuali generati da persone reali.

InstructLab è accessibile a sviluppatori ed esperti di dominio che potrebbero non avere le competenze necessarie in materia di data science normalmente richieste per il fine tuning dei LLM. La metodologia InstructLab consente ai team di aggiungere dati o competenze particolarmente adatte ai requisiti degli scenari di utilizzo aziendali al modello scelto, per un addestramento collaborativo che consente un time to value più rapido.

Addestramento e distribuzione in qualsiasi ambiente

Red Hat Enterprise Linux AI aiuta le organizzazioni ad accelerare il processo di passaggio da deployment Proof of Concept a deployment di produzione basati su server, fornendo tutti gli strumenti necessari e la possibilità di addestrare, ottimizzare e distribuire questi modelli dove si trovano i dati, in qualunque ambiente di cloud ibrido. I modelli distribuiti possono quindi essere utilizzati da vari servizi e applicazioni all'interno dell'azienda.



Quando le organizzazioni sono pronte, Red Hat Enterprise Linux AI fornisce anche un accesso agevolato a [Red Hat OpenShift® AI](#), per l'addestramento, l'ottimizzazione e la fornitura di questi modelli su vasta scala in un ambiente cluster distribuito utilizzando gli stessi modelli Granite e l'approccio InstructLab utilizzato nella distribuzione di Red Hat Enterprise Linux AI.

Caratteristiche e vantaggi

| Caratteristiche | Vantaggi |
|---|--|
| Modelli di codice e linguaggio Granite interamente supportati in open source con licenza Apache 2.0 | I LLM open source e trasparenti, insieme a dati di addestramento apertamente accessibili, migliorano la trasparenza dei dati e rispondono ai problemi etici relativi al contenuto e alle sorgenti dei dati, riducendo così il rischio aziendale complessivo. |

| Caratteristiche | Vantaggi |
|--|--|
| Indennizzi sulla proprietà intellettuale per i modelli Granite | Gli indennizzi per i modelli Granite all'interno di Red Hat Enterprise Linux AI rispecchiano la forte fiducia di Red Hat e IBM nello sviluppo e nei test rigorosi di questi modelli. Questi indennizzi forniscono ai clienti una maggiore sicurezza, consentendo loro di utilizzare i modelli Granite con maggiore fiducia nell'impegno di Red Hat per il loro successo. |
| Strumenti di allineamento dei LLM di InstructLab per il fine tuning di modelli scalabile e accessibile | InstructLab offre un metodo accessibile per il fine tuning dei LLM, riducendo la necessità di una conoscenza approfondita di data science e consentendo a vari ruoli all'interno dell'organizzazione di dare il proprio contributo. Questo consente alla tua azienda di adottare l'intelligenza artificiale generativa, accelerando il time to value e massimizzando il ritorno sull'investimento. |
| Istanze di runtime del modello ottimizzate e di avvio | Red Hat Enterprise Linux AI viene fornito come immagine del container di avvio, un metodo di distribuzione chiamato modalità immagine per Red Hat Enterprise Linux. Questa tecnologia riduce la complessità di installazione, configurazione e aggiornamento, consentendo un semplice processo di configurazione e gestione delle modifiche. |
| Dipendenze dei pacchetti dell'IA gen e driver software per l'hardware di intelligenza artificiale | Inizia subito a utilizzare l'IA generativa con un set completo di strumenti, tra cui pacchetti e driver essenziali come i driver PyTorch, vLLM e NVIDIA, per assicurarti di poter affrontare gli scenari di utilizzo dell'IA gen fin dall'inizio. |



Informazioni su Red Hat

Red Hat è leader mondiale nella fornitura di soluzioni software open source. Con un approccio che si avvale della collaborazione delle community, distribuisce tecnologie come Kubernetes, container, Linux e cloud ibrido caratterizzate da affidabilità e prestazioni elevate. Red Hat consente di sviluppare applicazioni cloud native, integrare applicazioni IT nuove ed esistenti e automatizzare e gestire ambienti complessi. [Considerata un partner affidabile dalle aziende della classifica Fortune 500](#), Red Hat fornisce [pluripremiati](#) servizi di consulenza, formazione e assistenza, che portano i vantaggi dell'innovazione open source in qualsiasi settore. Red Hat è l'elemento catalizzatore in una rete globale di aziende, partner e community, e permette alle organizzazioni di crescere, evolversi e prepararsi a un futuro digitale.

 facebook.com/RedHatItaly
 twitter.com/RedHatItaly
 linkedin.com/company/red-hat

ITALIA
it.redhat.com
italy@redhat.com

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

it.redhat.com
#1384734_0924

Copyright © 2024 Red Hat, Inc. Red Hat, il logo Red Hat e OpenShift sono marchi commerciali registrati di proprietà di Red Hat, Inc. o delle società da essa controllate con sede negli Stati Uniti e in altri Paesi. Linux® è un marchio registrato di proprietà di Linus Torvalds depositato negli Stati Uniti e in altri Paesi.