

Crea su una piattaforma di IA integrata con Red Hat AI Enterprise

La rapida ascesa dell'IA

L'IA è già stata adottata diffusamente, e ciò ha comportato che:

- ▶ Il 55% delle organizzazioni utilizza l'IA gen.¹
- ▶ Oltre il 50% delle organizzazioni esegue proof of concept o implementa scenari di utilizzo selezionati per l'Agentic AI.²
- ▶ Secondo le stime, il 30% delle applicazioni basate sull'IA utilizzerà l'Agentic AI entro la fine del 2026.²

L'evoluzione delle opportunità offerte dall'IA crea una complessità maggiore

L'IA ha offerto alle aziende l'opportunità di ridefinire il proprio modo di lavorare e competere. È diventata rapidamente un elemento chiave nel panorama aziendale odierno; tuttavia, nonostante la sua crescente diffusione, molte aziende devono affrontare sfide che ostacolano la piena valorizzazione dell'IA.

Alcune di esse sono l'aumento dei costi dei modelli, le complesse esigenze di personalizzazione, i rigidi vincoli di deployment e la necessità di stare al passo con il ritmo incalzante dell'innovazione.

Man mano che le aziende iniziano a concentrarsi sempre più sull'innovazione dell'IA gen, sui flussi di lavoro dell'Agentic AI e sull'IA sovrana, queste sfide diventano sempre più complesse da affrontare. Si rende quindi necessario un passaggio strategico verso piattaforme di IA che consentano alle aziende di ridurre i costi di inferenza, semplificare la scalabilità e il monitoraggio, creare un'IA sovrana on premise e adattarsi rapidamente ai cambiamenti.

Affronta la complessità dell'IA con piattaforme e strumenti destinati alle aziende

Che si tratti di aziende all'inizio del loro percorso verso l'IA o che stiano ampliando iniziative già consolidate all'interno dell'organizzazione, Red Hat® AI può aiutarle ad accelerare l'innovazione dell'IA e a ridurre i costi operativi di sviluppo e distribuzione di soluzioni di IA in ambienti ibridi e multicloud.

Red Hat® AI fornisce soluzioni convenienti grazie a modelli ottimizzati e un'inferenza efficiente, un'integrazione semplificata con i dati privati e una distribuzione rapida dei flussi di lavoro dell'Agentic AI su una piattaforma scalabile e flessibile. Consente alle organizzazioni di gestire e monitorare il ciclo di vita dei modelli di IA predittiva e IA gen in modo scalabile, dai deployment a server singolo alle piattaforme distribuite con scalabilità elevata.

A partire da comprovate tecnologie open source, Red Hat AI fornisce supporto per un'ampia gamma di modelli di IA open source per aiutare le aziende ad accelerare l'innovazione e a semplificare l'accesso a strumenti e tecnologie di valore.

A questo si aggiunge un ecosistema di partner per l'IA che offre prodotti e servizi testati e certificati incentrati su prestazioni, stabilità e supporto per acceleratori in tutte le infrastrutture per aiutare le aziende a risolvere le principali sfide presentate dall'IA dal punto di vista aziendale e tecnico.

¹ Presentazione di IDC Tech Buyer. "Tendenze di adozione iniziale degli scenari di utilizzo nel settore dell'IA generativa a livello mondiale, 2025: Riepilogo." Documento nr. US53280825, ottobre 2025. (è richiesto l'acquisto)

² Sondaggio IDC. "L'impatto dell'Agentic AI sulle strategie per l'infrastruttura digitale." Documento nr. US53418526, ottobre 2025. (è richiesto l'acquisto)

Red Hat AI offre:

- ▶ Monitoraggio e gestione completi del ciclo di vita dei modelli di IA.
- ▶ La flessibilità di utilizzare qualsiasi modello di IA su qualsiasi acceleratore hardware.
- ▶ Una raccolta convalidata di modelli di IA ottimizzati.
- ▶ Inferenza flessibile e conveniente in ambienti ibridi o multicloud.
- ▶ Integrazione affidabile nelle risorse dati private delle aziende.
- ▶ Distribuzione semplificata dei flussi di lavoro dell'Agentic AI.
- ▶ La possibilità di scalare secondo necessità con un utilizzo efficiente delle risorse.

Red Hat AI Enterprise è una piattaforma di IA unificata che fa parte del portfolio Red Hat AI per aiutare le aziende a risolvere le sfide legate all'adozione dell'IA oggi.

6 motivi per scegliere Red Hat AI Enterprise come piattaforma di IA

Red Hat AI Enterprise offre tutti gli strumenti e le funzionalità che occorrono per supportare le aziende in qualsiasi fase del loro percorso. Consente loro di sviluppare e distribuire modelli, agenti e applicazioni di IA efficienti e convenienti in ambienti ibridi e multicloud su una piattaforma di IA unificata.

Questa piattaforma offre alle aziende un ambiente di sviluppo pronto all'uso e funzionalità di IA essenziali, tra cui l'ottimizzazione dei modelli, l'inferenza ad alte prestazioni, la gestione dei flussi di lavoro dell'Agentic AI e la flessibilità di supportare qualsiasi modello, utilizzare qualsiasi hardware e distribuire ovunque per soddisfare i requisiti in continua evoluzione riguardanti la collocazione dei dati.

Semplifica i deployment dell'IA con un'inferenza flessibile ed efficiente

Con l'accelerare dell'innovazione dell'IA, anche i deployment si fanno sempre più complessi; ciò può comportare un utilizzo inefficiente delle risorse, un aumento dei costi operativi e una riduzione della qualità dei risultati.

Red Hat AI Enterprise offre un'inferenza dei modelli ottimizzata in diversi ambienti, inclusi ambienti on premise, ibridi e multicloud o all'edge di una rete. Ciò consente alle aziende di semplificare i deployment dei modelli utilizzando meno risorse, riducendo i costi di inferenza e mantenendo la giusta precisione.

Il segreto è un runtime ottimizzato fornito da vLLM, un server di inferenza che accelera l'output delle applicazioni basate sull'IA gen sfruttando al meglio la memoria GPU per ottimizzare la larghezza di banda e la latenza. Inoltre, sono incluse anche tecniche di compressione avanzate che contribuiscono a ottimizzare i costi riducendo le dimensioni e i requisiti computazionali di un modello.

Quando un'organizzazione è pronta per scalare, passando da pochi modelli a decine e da decine di utenti a centinaia, la piattaforma offre la possibilità di utilizzare llm-d. Questa tecnologia fornisce un framework di inferenza distribuita per prestazioni convenienti e prevedibili in modo scalabile per i modelli di IA gen a partire dalle innovazioni principali dei vLLM e offrendo prestazioni migliori con il crescere dei volumi.

Migliora la precisione dei modelli di IA con dati aziendali privati

Nonostante le funzionalità dei modelli di IA gen disponibili al pubblico, le aziende hanno bisogno di modelli di IA addestrati in base ai loro dati univoci per affrontare gli scenari di utilizzo specifici del dominio.

Red Hat AI Enterprise offre un'esperienza semplificata e coerente per le aziende, per supportare la personalizzazione efficiente dei modelli e aiutare ingegneri e data scientist di IA a migliorare l'accuratezza e la pertinenza delle risposte dei modelli. Consente alle aziende di connettere i modelli ai dati aziendali per gli scenari di utilizzo della Retrieval augmented generation (RAG).

Questa piattaforma supporta anche il fine tuning e l'apprendimento continuo e per rinforzo oltre a offrire un approccio flessibile e modulare per l'elaborazione e l'analisi dei documenti, la generazione di dati sintetici nonché la valutazione e le attività dei modelli.

Accelera l'innovazione con i flussi di lavoro dell'Agentic AI

La tecnologia relativa all'IA si è evoluta a pieno ritmo in un periodo di tempo relativamente breve. A partire dall'IA predittiva, che ha aperto molti nuovi scenari di utilizzo, si è passati alla diffusione dei modelli di IA gen; oggi, l'attenzione è rivolta verso i flussi di lavoro basati sull'Agentic AI, che rappresentano l'evoluzione successiva in grado di ampliare ulteriormente le strategie di IA e offrire ancora più valore aggiunto alle imprese.

Red Hat AI Enterprise offre una piattaforma agile e stabile che consente alle aziende di offrire valore aggiunto grazie all'Agentic AI. Questo risultato è possibile attraverso un'interfaccia di programmazione delle applicazioni (API) unificata, componenti in dotazione che semplificano l'assemblaggio dei flussi di lavoro dell'Agentic AI per gli scenari di utilizzo aziendali, esperienze utente dedicate e una base flessibile e scalabile che supporta il deployment e la gestione dei sistemi di Agentic AI.

La piattaforma fornisce anche il supporto per il Model Context Protocol (MCP), un componente fondamentale per il deployment agentic che funge da traduttore standardizzato tra un'ampia gamma di strumenti e funzionalità e i modelli linguistici di grandi dimensioni (LLM).

Estendi l'IA nel cloud ibrido in modo flessibile e coerente

In molti settori, le organizzazioni sono riuscite a rilevare il valore dell'IA già nei primi scenari di utilizzo. Tuttavia, sono ancora troppe le aziende che faticano a diffondere tale valore in tutta l'organizzazione mantenendo al contempo l'efficienza in termini di costi e il rispetto della conformità normativa, della privacy e della sicurezza. Ciò è particolarmente complicato nei casi di ambienti IT diversi, come gli ambienti ibridi e multicloud o i deployment dell'IA all'edge di una rete.

Red Hat AI Enterprise fornisce il controllo e la coerenza necessari per addestrare, ottimizzare, distribuire, gestire e rendere scalabili i modelli di IA predittiva e di IA gen nel modo più adatto possibile alla strategia relativa al carico di lavoro IA di un'organizzazione, in qualsiasi ambiente IT o hardware.

Questa piattaforma offre flessibilità perché è in grado di supportare qualsiasi modello di IA su qualsiasi combinazione tra hardware, produttori di apparecchiature originali (OEM), provider di servizi cloud e datacenter. Combinando tutto questo con una suite completa di funzionalità di osservabilità e monitoraggio, le aziende possono ottimizzare le prestazioni, controllare i costi, mantenere la governance e concentrarsi sulla sicurezza mentre evolvono per adattarsi a esigenze aziendali sempre diverse.

Mantieni la sovranità dei dati con deployment di IA on premise

In tutto il mondo, continua l'introduzione di norme governative sull'utilizzo dell'IA e sulle modalità di archiviazione ed elaborazione dei dati. Per questo, molte aziende hanno iniziato a cercare di capire come creare deployment di IA on premise per mantenere la sovranità e la privacy.

Archiviando ed elaborando i propri dati sensibili e privati on premise, le aziende possono non solo conformarsi alle tante normative in continua evoluzione, ma anche salvaguardare la privacy di tali dati con una superficie di attacco più ridotta e una maggiore attenzione alla sicurezza.

Red Hat offre un approccio completo e stratificato alla sicurezza dell'IA, consentendo alle organizzazioni di adottare in tutta tranquillità innovazioni dell'IA. Red Hat AI Enterprise offre questo approccio alla sicurezza su una base di cloud ibrido e open source affidabile, per affrontare le sfide specifiche legate alla sicurezza del ciclo di vita dell'IA.

Favorisce una sovranità reale supportando qualsiasi modello, hardware e condivisione di GPU incentrata sulla sicurezza in tutti gli ambienti cloud e edge, compresi gli ambienti isolati. Red Hat lavora a stretto contatto con i provider di cloud sovrano per consentire alle organizzazioni di distribuire l'IA in ambienti cloud privati e incentrati sulla sicurezza.

Red Hat AI Enterprise include anche strumenti che supportano l'affidabilità verificabile tramite l'esplicabilità dei modelli, l'equità e l'applicazione della policy di output, nonché la tracciabilità del ciclo di vita con pipeline riproducibili e la preparazione agli audit.

Sviluppa l'IA su una piattaforma Kubernetes affidabile

Red Hat AI Enterprise offre questo set completo di funzionalità di IA come parte di uno stack di IA end to end integrato e pronto all'uso costruito sulla base affidabile di Kubernetes.

Tutto questo garantisce un'esperienza conosciuta e coerente nei datacenter on premise, negli ambienti ibridi o multcloud e nei deployment di IA all'edge alle aziende che già utilizzano Kubernetes e la containerizzazione. Le aziende possono creare, addestrare, distribuire e gestire applicazioni moderne, cloud native e basate sull'IA su un'unica piattaforma centralizzata utilizzando gli stessi strumenti, framework e competenze già noti ai loro team.

Scopri di più sui vantaggi di Red Hat AI Enterprise

[Scopri di più](#) su come Red Hat AI Enterprise aiuta le organizzazioni a distribuire l'IA in modo efficiente e conveniente nel cloud ibrido con una piattaforma unificata che supporta ogni fase del percorso di adozione dell'IA.



Informazioni su Red Hat

Red Hat è leader mondiale nella fornitura di soluzioni software open source. Con un approccio basato sul concetto di community, distribuisce tecnologie affidabili e altamente performanti come Linux, cloud ibrido, container e Kubernetes. Red Hat consente di sviluppare applicazioni cloud native, integrare applicazioni IT nuove ed esistenti, nonché 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.