

Definizione della strategia aziendale per l'IA: una guida introduttiva



Red Hat

Introduzione

Sempre più organizzazioni riconoscono le opportunità che l'intelligenza artificiale (IA) offre in ogni campo di attività.

Dall'interazione con i clienti al supporto e alla vendita, passando per l'infrastruttura IT, i processi, lo sviluppo del codice e la distribuzione delle soluzioni, la diffusione e lo slancio dell'IA in ogni scenario di utilizzo e settore sembrano inarrestabili.

Secondo IDC, entro il 2027 il mercato avrà superato i 423 miliardi di dollari, con un tasso di crescita annuale composto (CAGR) del 26,9% su cinque anni. Le iniziative di IA delle aziende punteranno soprattutto a migliorare l'efficienza operativa, l'esperienza dei clienti e la produttività.¹

Presi nel vortice di questa rapida evoluzione, i responsabili sono gravati dalla pressione a identificare, selezionare, creare ed erogare le soluzioni IA più idonee per offrire alle proprie organizzazioni un vantaggio competitivo. Tuttavia, la velocità di innovazione dell'IA e la capacità della maggior parte delle aziende di migliorare il proprio livello di maturità rispetto all'IA non sono allineate. Sfruttare al meglio il pieno valore dell'IA risulta quindi ancora complesso, e spesso genera più domande che risposte.



Questo ebook prova a rispondere a queste domande, che provengano da chi ha appena avviato il percorso verso l'IA, da chi cerca di comprendere quale impatto può avere sulla sua azienda o da chi cerca di capire come rendere scalabili le implementazioni di IA già esistenti.

¹ IDC FutureScape Webcast, "[Worldwide Artificial Intelligence and Automation 2024 Predictions](#)", documento n. US51901124, marzo 2024.

Tipologie di IA

Per sfruttare al meglio l'IA occorre saperne il più possibile, a partire dai due tipi di IA più utilizzati al momento dalle aziende.

IA predittiva: utilizzando dati storici, l'IA predittiva aiuta le organizzazioni a identificare gli schemi e a migliorare il processo decisionale sugli eventi futuri. I modelli predittivi rendono più efficienti le applicazioni che prevedono la domanda, definiscono le attività di manutenzione preventiva o pianificano le operazioni. Basata su tecnologie di data science e machine learning (ML) consolidate, l'IA predittiva migliora le proprie funzionalità all'aumentare dei dati elaborati.



IA generativa: l'IA generativa, alimentata da modelli di deep learning trasformativi, è in grado di creare nuovi contenuti in forma di testo, immagini e codice. Si rivela particolarmente utile per applicazioni chatbot, per la generazione di contenuti automatizzata e per gli strumenti di creatività. Modelli come GTP (Generative Pre-trained Transformer) hanno rivoluzionato i campi dell'elaborazione del linguaggio naturale e della creatività, producendo testi e immagini simili a quelli prodotti dall'intelligenza umana.

I vantaggi dell'adozione dell'IA

Benché le potenzialità effettive dell'IA debbano ancora emergere appieno, comprendere quali siano i vantaggi che questa tecnologia in rapida evoluzione offra già alle aziende di ogni dimensione e settore è un ottimo punto di partenza per capire dove può essere applicata nella tua organizzazione.

Esamina i vantaggi offerti dall'IA elencati di seguito per capire come possono contribuire alla tua attività.



101011011

Volume dei dati. La crescita esponenziale dei dati complica la loro gestione e l'estrapolazione di informazioni fruibili. Capace di elaborare e analizzare in breve tempo set di dati di grandi dimensioni, l'IA può rivelare informazioni e tendenze altrimenti difficili da identificare manualmente.



Bar chart with gear icon

Inefficienza operativa. Processi inefficienti e colli di bottiglia possono rallentare la produttività e per eliminare questi ostacoli occorrono tempo e impegno. L'automazione basata sull'IA può contribuire a ottimizzare le operazioni, riducendo gli errori e migliorando l'efficienza dei processi. Ne sono esempi le applicazioni che elaborano automaticamente i verbali degli incontri indicando le azioni da svolgere e i successivi passaggi da compiere, o che velocizzano la creazione di immagini e video destinati a siti web o social media.



Person silhouette icon

Aspettative dei clienti. I clienti si aspettano esperienze personalizzate e senza intoppi. L'IA è in grado di migliorare e personalizzare il servizio offerto grazie all'analisi dei dati dei clienti, fornendo raccomandazioni e interazioni su misura.



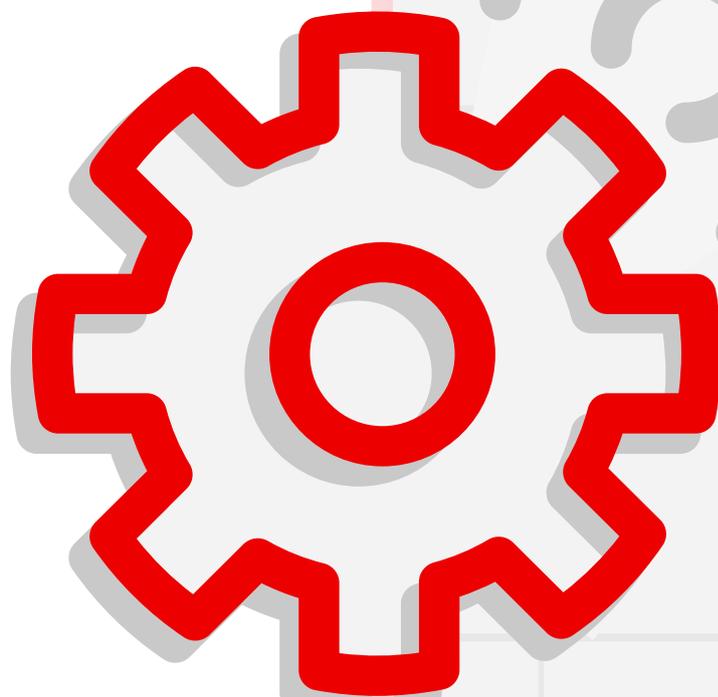
Lightbulb icon

Competitività del mercato. L'innovazione continua è imprescindibile per rimanere competitivi in un mercato in costante cambiamento. L'IA può aiutare le aziende ad adattarsi ai cambiamenti del mercato restando competitive. L'IA generativa può aiutare a perfezionare l'approccio se utilizzata come consulente per la condivisione delle idee o durante la preparazione di un incontro importante.

L'avvento dell'intelligenza artificiale

In evoluzione ormai da decenni, l'IA apre nuove opportunità di progresso in settori come quello sanitario, finanziario e manifatturiero.

La recente diffusione dell'IA generativa, tuttavia, ha catalizzato l'attenzione per la sua capacità di produrre testi, immagini realistiche e codice software simili a quelli prodotti dal cervello umano. A differenza dell'IA convenzionale, utilizzata per automatizzare le attività o analizzare i dati, l'IA generativa apre le porte alla risoluzione creativa dei problemi e alla creazione di contenuti avanzati.



Tipologie di modelli di IA che accelerano l'innovazione

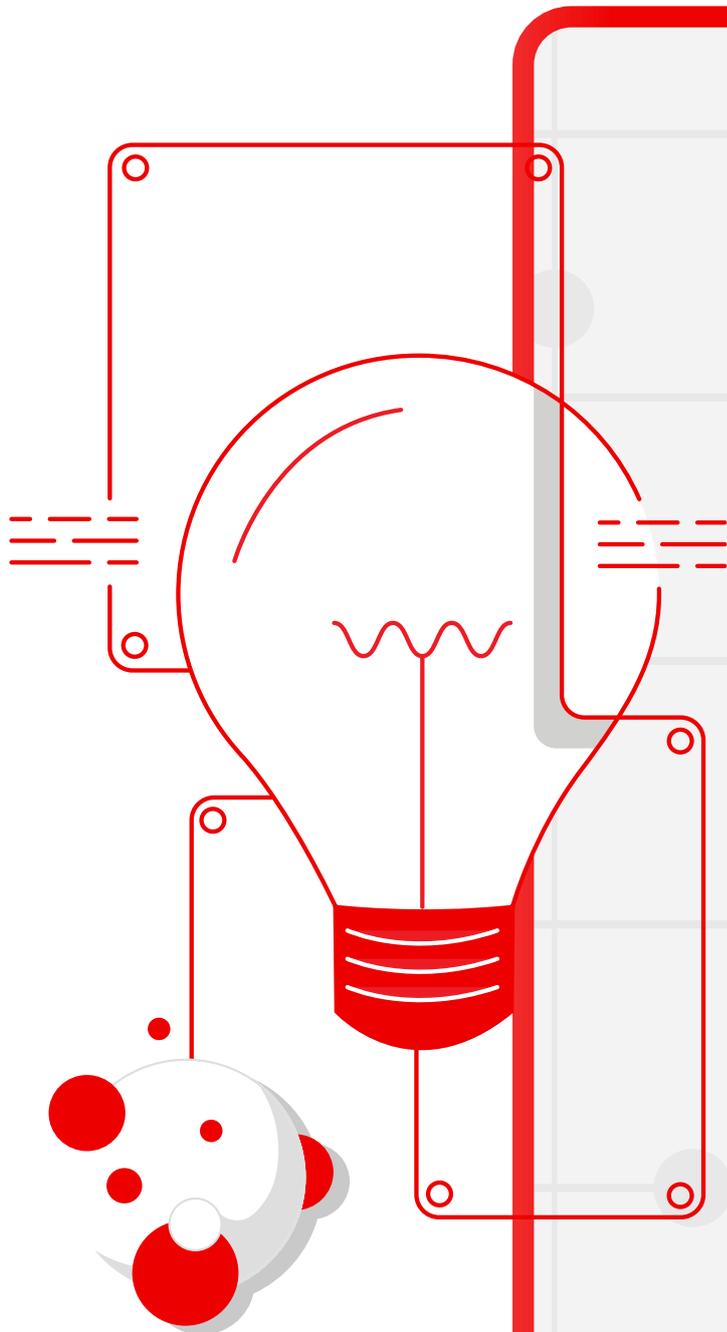
Tra i modelli di IA responsabili della crescita esponenziale dell'IA generativa troviamo i [modello linguistici di grandi dimensioni](#) (LLM, Large Language Model) e i modelli a diffusione stabile. I modelli LLM, di cui GPT è un esempio, vengono addestrati con grandi set di dati e sono in grado di comprendere e di generare il linguaggio naturale. Questa capacità li rende preziosi per l'automazione delle attività di assistenza ai clienti, di creazione di testi di marketing e molto altro. I modelli a diffusione stabile possono invece creare immagini iperrealistiche, alimentando l'innovazione nei settori dell'intrattenimento, del marketing e altri.

Tendenze emergenti

L'IA multimodale, verso cui si rivolge oggi l'interesse delle aziende, è caratterizzata da un unico modello in cui convergono capacità di elaborazione di testi, immagini e dati per un'offerta di soluzioni più versatili. Restare al passo delle nuove tendenze è cruciale per sfruttare appieno le potenzialità dell'IA negli ambienti aziendali.

Open source: un punto di partenza per l'innovazione dell'IA

Profondamente radicata nell'open source, la [strategia per l'IA di Red Hat](#) può aiutare le aziende a migliorare l'IA generativa con trasparenza, affidabilità e costi ridotti. Utilizzando le piattaforme di [cloud ibrido](#) open source di Red Hat, le organizzazioni possono innovare liberamente, mantenendo il controllo sulle proprie soluzioni di IA.



[Scopri di più sugli LLM e sul loro funzionamento](#)

Contribuisci allo sviluppo degli LLM con l'open source

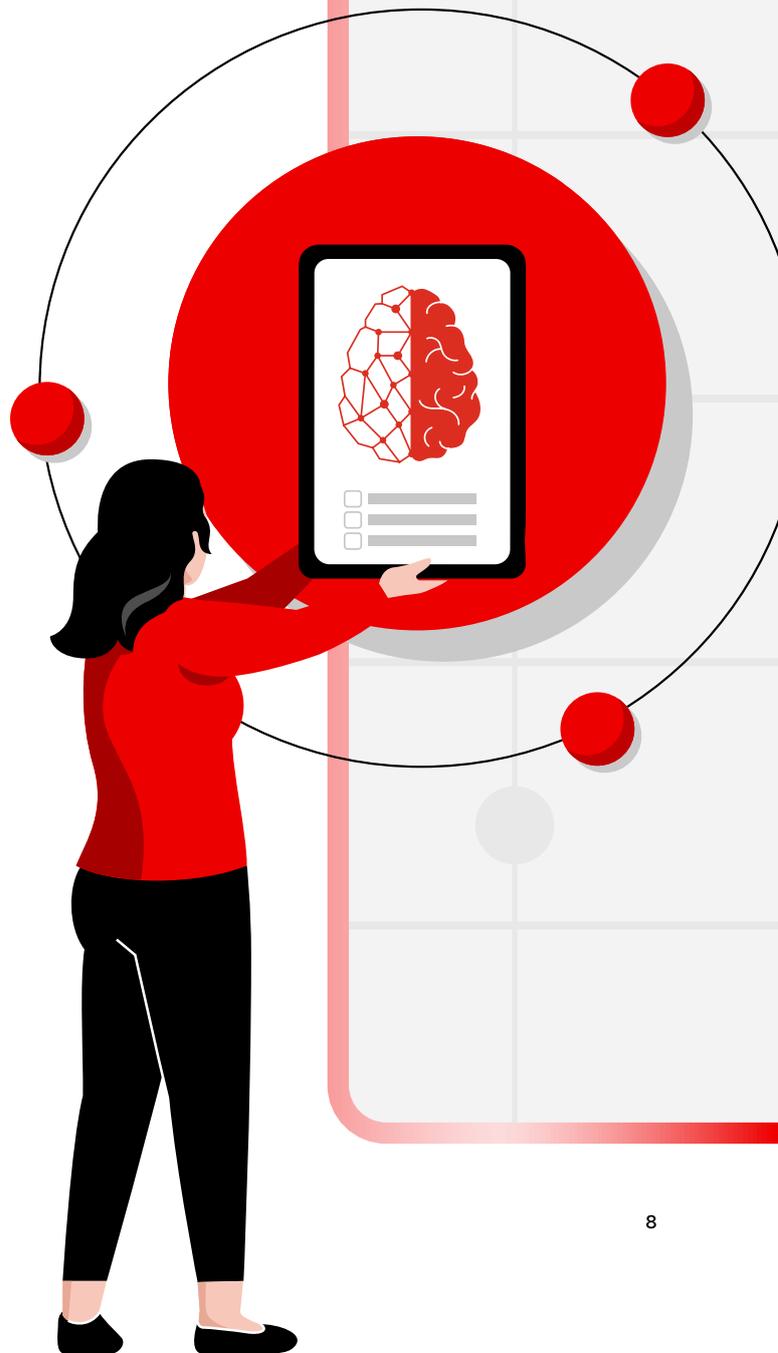
Mentre l'IA generativa stimola il cambiamento in quasi ogni aspetto delle attività aziendali, dalle modalità di sviluppo del software a quelle di comunicazione, non è raro che i modelli (sia LLM che di altro tipo) che alimentano una delle funzionalità di IA generativa restino sotto il controllo del fornitore del servizio. Di conseguenza, un'azienda che non dispone di competenze qualificate e di tempo e budget sufficienti avrà difficoltà a valutare le capacità di un servizio di IA generativa.

La mancanza di visibilità sui set di dati alla base del modello o di informazioni su come il modello utilizza i dati espone l'azienda a potenziali rischi rispetto ai contenuti che l'IA genera, ad esempio in termini di proprietà intellettuale, sia nel caso in cui il modello generativo sia stato addestrato con un codice sorgente tutelato da diritto di autore, ma anche se il codice stesso che il modello genera è parte del codice tutelato. A questa e a molte altre questioni simili non c'è ancora una risposta esaustiva. Sapendo che le conseguenze possono essere serie, molte aziende rivolgono il proprio interesse all'IA open source.

L'approccio di Red Hat all'IA si fonda sull'open source, così come il supporto che forniamo ai modelli open source della famiglia di modelli fondativi [IBM Granite](#).

Le soluzioni IA di Red Hat contribuiscono direttamente allo sviluppo di modelli di IA con [InstructLab](#), una soluzione promossa dalla community per migliorare i modelli LLM.

[Esplora InstructLab su GitHub](#)



Scegli il modello di IA più adeguato

I diversi modelli di IA possono essere applicati a diversi scenari di utilizzo.

I modelli di IA predittiva e IA generativa possono essere utilizzati insieme in un'unica applicazione o servizio; ciascuno di essi avrà costi diversi e offrirà vantaggi specifici, ma tutti contribuiranno a ridurre i tempi di rilascio per la Proof of Concept iniziale. I modelli per la segmentazione e il riconoscimento delle immagini o per la trasformazione della voce in testo sono esempi diffusi e molto funzionali, ma è fondamentale saper valutare quale sia il più adatto al proprio contesto di utilizzo.

I modelli fondativi, addestrati con grandi volumi di dati, offrono grande versatilità ma la loro dimensione può accrescere i costi, complicare la gestione e aumentare la complessità. Per queste ragioni non sono adatti a tutti gli scenari di utilizzo.

Sempre parte della famiglia dell'IA generativa, i modelli più piccoli e ben ottimizzati costituiscono una soluzione migliore laddove è preferibile perfezionare un modello esistente in base alle proprie esigenze. Un modello predefinito può essere una valida scelta, perché è già disponibile e facilmente integrabile nei sistemi esistenti. Gli LLM sono esempi di questo ultimo tipo, in quanto strumenti potenti e già addestrati con grandi volumi di dati.

In caso di requisiti aziendali specifici, timori sulla privacy dei dati o esigenze di controllo sul comportamento del modello, potrebbe essere necessario creare un modello personalizzato self-hosted.

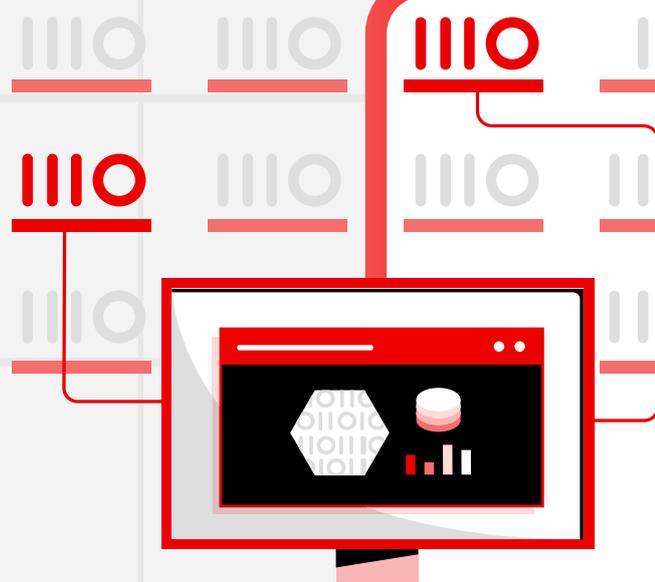
È meglio creare un modello o ottimizzarne uno predefinito?

Creare un modello di IA da zero può rivelarsi un compito gravoso. Innanzitutto occorre raccogliere e preparare grandi set di dati pertinenti rispetto allo scenario di utilizzo aziendale previsto. Occorre poi scegliere l'algoritmo adeguato e addestrarlo con i dati acquisiti. Servono per questo grande esperienza e potenza di elaborazione, il che rende il processo dispendioso in termini di tempo e di risorse. Costruire un modello personalizzato, che sia fondativo o tradizionale, garantisce senz'altro una soluzione su misura, ma non è sempre la strada migliore.

È vero anche che l'ottimizzazione di un modello fondativo implica l'adattamento di un modello preaddestrato a determinati requisiti specifici. Un approccio diffuso è il transfer learning, in cui un modello addestrato su un set di dati di grandi dimensioni viene poi riaddestrato con un set di dati più piccolo e specifico per un determinato dominio. Con questo modello le conoscenze generiche apprese durante l'addestramento iniziale non vanno perse ma vengono adattate alle particolarità dei dati specifici.

Ottimizzazione del modello

Un altro approccio è il fine tuning, che prevede l'adattamento dei parametri del modello per migliorare le prestazioni di un'attività specifica. I parametri del modello sono le variabili di un modello selezionato che possono essere stimate applicando al modello dati specifici. Il fine tuning può comportare la modifica della velocità di apprendimento, dell'architettura del modello oppure l'addestramento intensivo di alcuni livelli del modello rispetto ad altri. Queste tecniche puntano a migliorare le conoscenze del modello, rendendolo più efficiente per gli scenari di utilizzo specifici dell'utente.



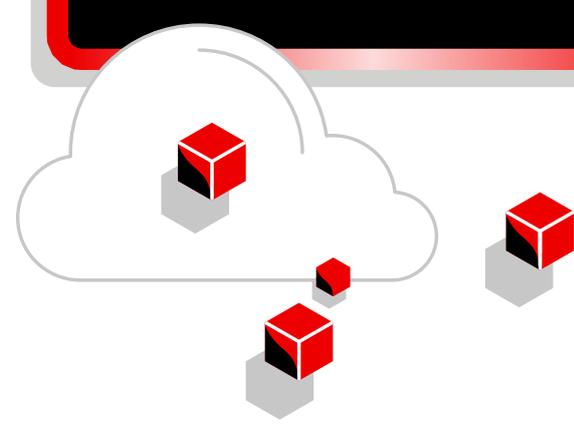
[InstructLab](#) usa il fine tuning con l'obiettivo di ridurre le competenze sull'IA necessarie e semplificare l'integrazione delle conoscenze aziendali nei modelli fondativi Granite esistenti.

Alternative all'ottimizzazione dei modelli

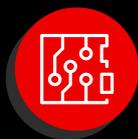
I ricercatori stanno studiando nuovi modi per accelerare e razionalizzare l'ottimizzazione dei modelli fondativi. Una delle tecniche più diffusa è la [RAG \(Retrieval-Augmented Generation\)](#) che consente di recuperare dati selezionati da una sorgente esterna nella quale sono state codificate conoscenze aggiuntive, ovvero il contesto.

La RAG utilizza uno o più database esterni (anche detti database vettore) che forniscono il contesto aggiuntivo per rispondere meglio alla domanda posta al modello di IA generativa. I sistemi Agentic AI costituiscono un altro approccio innovativo. Combinano una serie di agenti di IA generativa per interrogare sistemi esterni, ad esempio database interni, intranet aziendali o Internet, per fornire ai modelli di IA generativa le informazioni più accurate e aggiornati disponibili.

L'ultimo esempio è il [fine tuning del prompt](#), in cui i modelli di IA ricevono suggerimento o prompt front end, come parole extra o numeri generati dall'IA, che li indirizzano verso la decisione auspicata. Il risultato di una query RAG crea ulteriore contesto per il prompt; in questo caso l'ottimizzazione del prompt e la RAG funzionano in sinergia. Combinando il fine tuning di un LLM esistente, l'uso di tecniche RAG e il fine tuning del prompt, anche le organizzazioni con dati limitati possono personalizzare un modello fondativo affinché svolga attività specifiche.

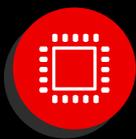


L'infrastruttura che supporta i modelli di IA è importante tanto quanto il modello stesso. Per svolgere attività differenti servono tipologie di hardware diverse.



CPU **(Central Processing Unit)**

Processori tradizionali che gestiscono attività di elaborazione generiche. Versatili, ma non sempre efficienti per i carichi di lavoro di IA di grandi dimensioni.



GPU **(Graphics Processing Unit)**

Processori specializzati progettati per gestire attività di elaborazione in parallelo, e quindi perfetti per l'addestramento di modelli di deep learning che richiedono l'elaborazione simultanea di grandi quantità di dati.



NPU **(Neural Processing Unit)**

Recente tipologia di processori progettati appositamente per le attività correlate all'IA, che offre più efficienza e velocità per alcuni tipi di modelli.

Il ruolo del cloud ibrido nell'adozione dell'IA in azienda

Gli ambienti di cloud ibrido sono fondamentali per l'adozione dell'IA. Combinando l'infrastruttura on premise e le risorse di cloud pubblico e privato, il cloud ibrido permette di distribuire e gestire i carichi di lavoro di IA quando e dove necessario, con grande flessibilità. È possibile ad esempio addestrare i modelli di IA con potenti GPU nel cloud e quindi distribuirli in locale o in un cloud privato se necessario per ragioni di sicurezza o conformità. Prima di adottare l'approccio al cloud ibrido è importante valutare la coerenza degli strumenti e della piattaforma selezionata.

L'approccio al cloud ibrido open source di Red Hat facilita l'integrazione dell'IA nei diversi ambienti, migliorando coerenza, scalabilità e flessibilità. Permette inoltre di gestire i carichi di lavoro IA in più ambienti cloud, on premise o all'edge della rete, di ottimizzare il posizionamento dei dati e di semplificarne il trasferimento, il che si traduce nell'adozione semplificata dell'IA a livello aziendale.

Comprendere i modelli, i dati e l'infrastruttura dell'IA permette di affrontare meglio le difficoltà della sua adozione e di sfruttarla al meglio.

Cosa serve per iniziare

Come per l'adozione di qualsiasi nuova tecnologia, la riuscita delle iniziative di IA di un'organizzazione dipende dal superamento di alcuni ostacoli.

Le considerazioni seguenti ti permettono di valutare se la tua organizzazione è pronta e di identificare le aree su cui concentrarti per accelerare l'adozione dell'IA.



Valuta la qualità e la disponibilità dei dati. Per la riuscita delle iniziative di IA è fondamentale poter accedere a dati pertinenti e di alta qualità. Essendo la qualità indispensabile per addestrare in modo accurato i modelli di IA, è importante valutare la completezza, l'accuratezza e la pertinenza delle informazioni disponibili.



Valuta l'infrastruttura tecnologica. L'infrastruttura esistente è in grado di supportare i carichi di lavoro di IA? Sono disponibili risorse di elaborazione, soluzioni di storage, automazione e capacità di rete ad alte prestazioni?



Identifica la necessità di ulteriori competenze. Valuta la disponibilità di esperti in IA nella tua organizzazione. In questo modo potrai analizzare le competenze esistenti e identificare l'eventuale necessità di formazione o di competenze specializzate.



Valuta l'allineamento strategico. Verifica che le iniziative di IA siano in linea con gli obiettivi e le strategie aziendali. I progetti di IA devono sostenere gli obiettivi strategici complessivi dell'organizzazione e garantire un valore aziendale misurabile.

Muovere i primi passi con l'IA

La velocità e la portata dell'adozione dell'IA a livello organizzativo dipendono da numerosi fattori, ma in linea di massima l'approccio migliore per qualsiasi progetto di modernizzazione tecnologica è l'implementazione graduale.

Gli otto passaggi seguenti possono aiutare la tua organizzazione ad avviare e a proseguire il percorso di adozione dell'IA.

1

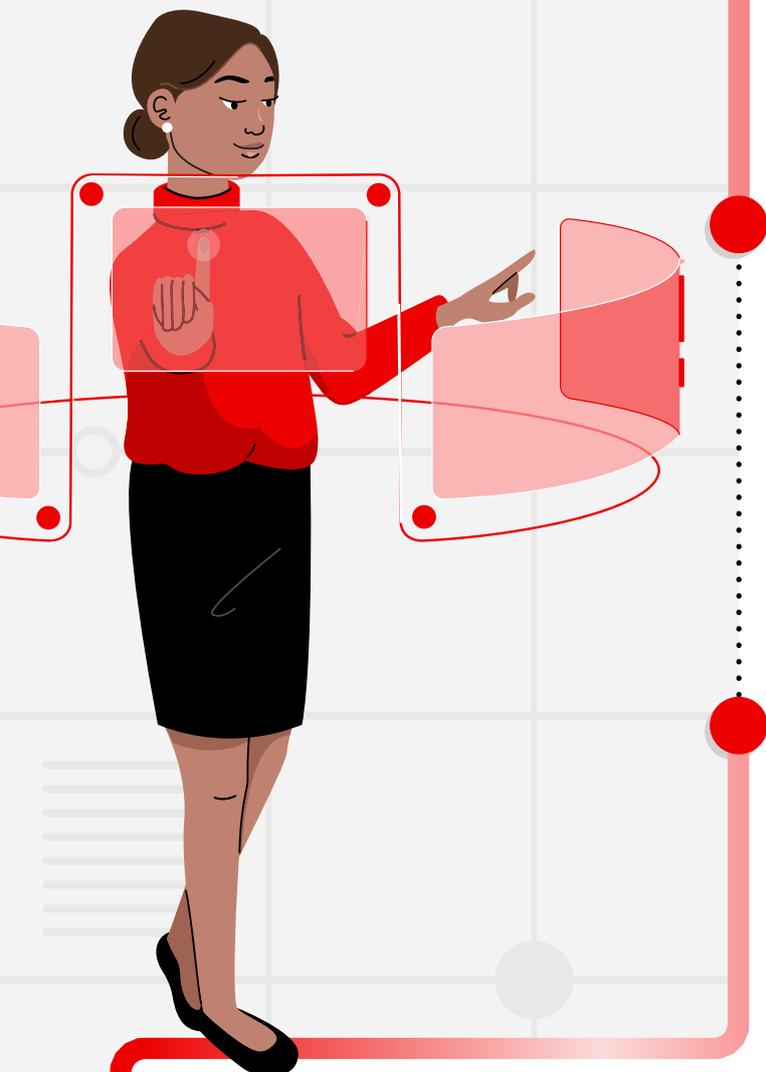
Valutazione di abilità e obiettivi

Il primo passo è la valutazione delle capacità, dell'infrastruttura e degli obiettivi strategici attuali dell'organizzazione. In questo modo è possibile stabilire se l'IA è in linea con gli obiettivi generali e identificare le aree potenziali a cui l'IA può aggiungere valore. Questa valutazione iniziale ti aiuta a stabilire qual è la direzione migliore del percorso di adozione dell'IA.

2

Identificazione degli scenari di utilizzo e dei team di IA

Identifica le opportunità che l'IA può offrire alla tua organizzazione. Organizza un team dedicato alle applicazioni IA che possa guidare l'iniziativa e che sia composto da membri di diverse funzioni come sviluppatori, esperti di dominio, data scientist e specialisti IT. Uno scenario di utilizzo ben specificato aiuterà ad adottare l'IA e a focalizzarsi sulle risorse.



3

Selezione del modello.

Scegli il modello di IA adeguato agli scenari di utilizzo identificati. Che si tratti di un LLM per l'IA generativa o di un modello predittivo per l'analisi dei dati, accertati che le capacità del modello siano adeguate agli obiettivi del tuo scenario di utilizzo. Dovrai tenere conto di fattori quali la complessità, la scalabilità e la compatibilità del modello con i sistemi esistenti.

4

Cicli di test e convalida

Definisci i criteri con cui valutare la riuscita dell'adozione dell'IA, ad esempio metriche prestazionali, percentuali di precisione o obiettivi aziendali. Stabilisci cicli di test e convalida per valutare costantemente l'efficienza del modello. I feedback regolari ottenuti da questi cicli contribuiscono all'ottimizzazione del modello e a mantenere la giusta direzione del percorso.

5

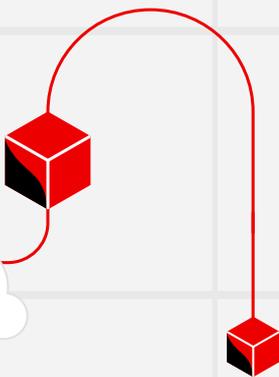
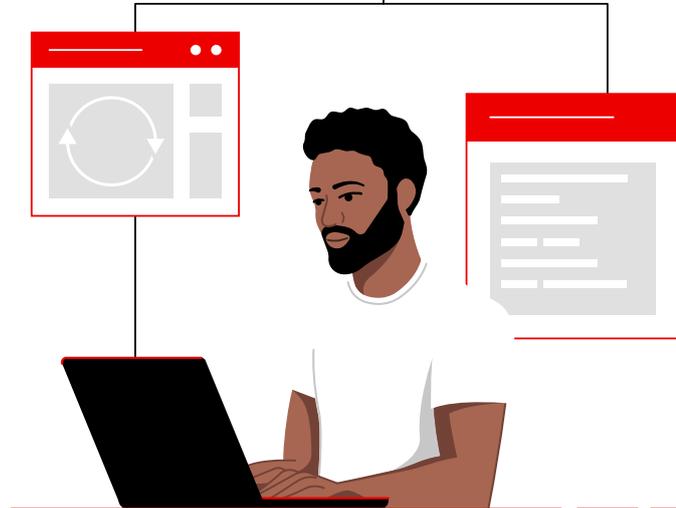
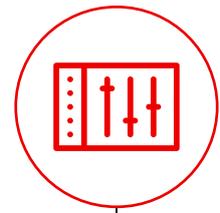
Ottimizzazione dei modelli

Il modello selezionato deve essere personalizzato utilizzando i dati dell'organizzazione. Questo processo implica l'alimentazione del modello con dati pertinenti per renderlo più preciso e pertinente allo scenario aziendale di utilizzo specifico. L'ottimizzazione ha lo scopo di ottenere un modello adatto al contesto e alle esigenze specifiche dell'organizzazione.

6

Addestramento con dati artificiali

Per perfezionare e addestrare ulteriormente i modelli di IA, valuta l'impiego di dati artificiali. Questo approccio utilizza metodi LLM insegnante-studente, e permette di generare dati di addestramento di alta qualità anche se i dati reali sono scarsi o riservati. I dati artificiali aiutano a migliorare la solidità e le prestazioni del modello, senza comprometterne la privacy.



7

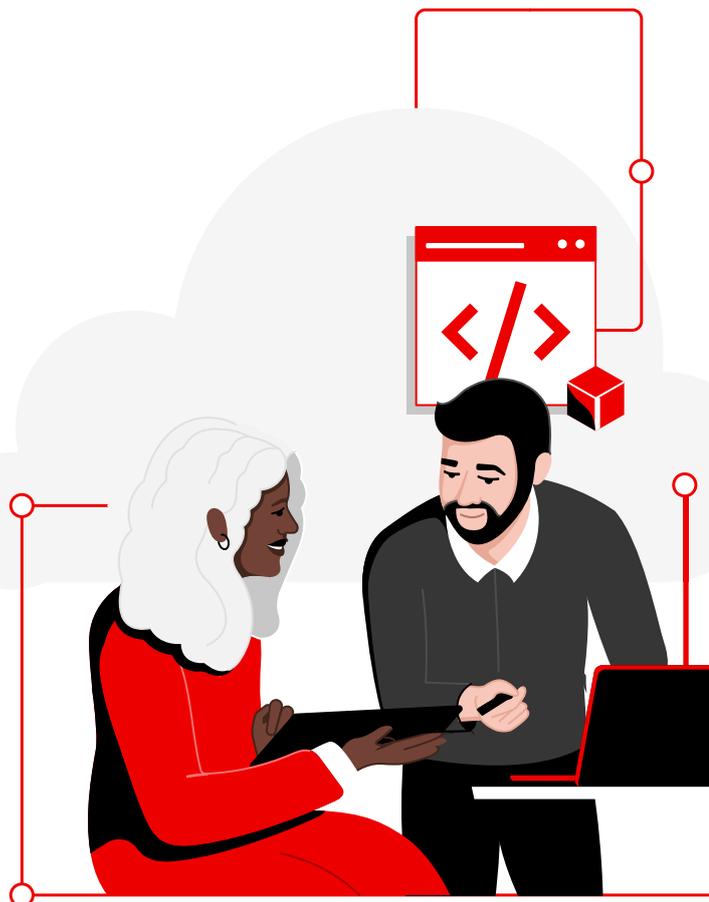
Monitoraggio delle deviazioni

Il monitoraggio delle deviazioni offre un sistema di monitoraggio generico basato sul contenuto. Rispetto alla gestione strutturata della configurazione, il monitoraggio delle deviazioni registra le modifiche apportate al contenuto nel file system locale. Aiuta a rilevare e a correggere le deviazioni, migliorando nel tempo la precisione del modello. Il monitoraggio continuo garantisce che il modello resti efficiente e pertinente anche al variare delle condizioni.

8

Contatta i nostri esperti

Se ritieni che l'esperienza con l'IA all'interno dell'azienda non sia ancora sufficiente, puoi chiedere il supporto degli esperti esterni di Red Hat® Consulting e ottenere molti vantaggi. Gli esperti di Red Hat possono guidarti attraverso gli aspetti più complessi dell'adozione dell'IA, fornendo informazioni preziose e l'accesso alla formazione pertinente. Con Red Hat Consulting puoi velocizzare il tuo percorso verso l'IA e aumentare le possibilità di riuscita.



L'adozione dell'IA si fonda sulla collaborazione

Creare un team interfunzionale che abbia capacità diversificate è essenziale per il successo dei progetti di IA generativa.²

- **Leader aziendali:** utilizzano la soluzione o usufruiscono dei vantaggi che offre.
- **Specialisti dell'IA:** ottimizzano, gestiscono e aggiornano i modelli di IA generativa.
- **Data scientist:** elaborano e forniscono dati corretti e neutri per l'addestramento dei modelli.
- **Responsabili dell'etica e della conformità:** garantiscono che le iniziative di IA generativa siano conformi alle normative.
- **Specialisti delle operazioni IT:** integrano le soluzioni nell'infrastruttura esistente e applicano i criteri di sicurezza.
- **Team e community di sviluppo:** devono essere coinvolti dall'inizio per collaborare, creare, condividere e migliorare gli strumenti open source, i framework e le procedure consigliate per l'adozione dell'IA. Solo così le modalità di utilizzo dell'IA saranno correlate al valore aziendale.

² Kearney, "[Standing up tiger teams to tackle generative AI complexity](#)", 15 novembre 2023.

Adozione e scalabilità con Red Hat

Red Hat AI assicura fiducia, possibilità di scelta e coerenza nel cloud ibrido, elementi in grado di accelerare l'adozione aziendale dell'IA.

Il portfolio Red Hat AI include Red Hat Enterprise Linux® AI, per i server Linux individuali, e Red Hat OpenShift® AI, per le piattaforme Kubernetes distribuite, e fornisce funzionalità integrate di machine learning. Entrambe le soluzioni sono basate su tecnologie e modelli open source e aiutano le organizzazioni ad essere sempre all'avanguardia nel campo dell'innovazione dell'IA, a scoprire nuove soluzioni più rapidamente e a democratizzare l'accesso agli strumenti e alle tecnologie più avanzati.

Il vasto ecosistema di partner Red Hat è in grado di rafforzare ulteriormente le funzionalità dell'IA. NVIDIA, ad esempio, l'azienda nota per aver diffuso le GPU, continua a collaborare con Red Hat per ampliare le potenzialità dell'IA con una piattaforma aziendale end to end ottimizzata per i carichi di lavoro di IA. Con NVIDIA, i clienti aziendali possono adottare con facilità il computing accelerato da GPU, per applicazioni di elaborazione ad alte prestazioni finalizzate all'uso dell'IA.

"Red Hat e NVIDIA vantano una collaborazione di lunga data. Red Hat Enterprise Linux AI testimonia il nostro impegno condiviso ad avvicinare l'intero stack di elaborazione e software agli sviluppatori e ai ricercatori che lavorano alla prossima ondata di tecnologie e applicazioni IA"³.

Justin Boitano, Vice President, Enterprise Products, NVIDIA

³ Comunicato stampa di Red Hat, "[Red Hat innova e rende accessibile l'IA generativa open source con Red Hat Enterprise Linux AI](#)", 7 maggio 2024.



Approfondimento su Red Hat Enterprise Linux AI

Red Hat Enterprise Linux AI è composto da quattro componenti di base:

1

Modelli Open Granite

Red Hat Enterprise Linux AI comprende dei modelli Granite di codice open source completamente tutelati da Red Hat. Questi modelli flessibili consentono di creare modelli linguistici personalizzati e utilizzarli in modalità pubblica o privata.

2

Allineamento dei modelli InstructLab

InstructLab è un progetto open source promosso da Red Hat e IBM per la personalizzazione dei modelli di IA tramite conoscenze specifiche e la generazione di dati artificiali da utilizzare per l'addestramento. È uno strumento a riga di comando integrato con un repository git che consente agli utenti di aggiungere competenze e addestrare i modelli con facilità.

3

Red Hat Enterprise Linux ottimizzato e avviabile per modelli Granite e InstructLab

I modelli Granite e gli strumenti InstructLab vengono eseguiti in un'immagine di Red Hat Enterprise Linux specializzata e ottimizzata per l'IA, compatibile con tutti gli ambienti hardware e cloud. Questa configurazione garantisce prestazioni efficienti con GPU di alta qualità, indispensabili per accelerare l'addestramento e il deployment dei modelli.

4

Supporto e indennizzi di livello enterprise

Le sottoscrizioni Red Hat Enterprise Linux AI includono il supporto di livello enterprise, un ciclo di vita del prodotto completo che inizia dal modello e dal software Granite 7B e indennizzi per gli IP forniti da Red Hat.

Red Hat Enterprise Linux AI aiuta a concretizzare le applicazioni dell'IA generativa

Alle organizzazioni che muovono i primi passi con l'IA generativa, Red Hat Enterprise Linux AI offre LLM e modelli linguistici di codice già predisposti in un unico ambiente di sviluppo e inferenza su server.

Questo ambiente unificato con modelli e strumenti facilita l'avvicinamento all'IA generativa e alla personalizzazione dei modelli utilizzando i dati aziendali, senza richiedere esperienza in IA o un'infrastruttura specializzata.

Completamente supportato e tutelato da Red Hat, Red Hat Enterprise Linux AI riduce i rischi. Inoltre, offre un approccio semplificato all'IA generativa progettato per essere più accessibile agli sviluppatori e agli esperti di dominio, promuovendone la collaborazione e riducendo il tempo necessario a concretizzare i risultati aziendali.

Perché scegliere Red Hat Enterprise Linux per l'IA?

LLM per l'azienda

Gli LLM di IBM Granite con licenza open source sono disponibili con la licenza Apache-2.0 e sono completamente supportati e tutelati da Red Hat.

Collaborazione della community

InstructLab semplifica la sperimentazione e l'ottimizzazione dell'allineamento dei modelli di IA generativa.

Scalabilità cloud native

La modalità immagine di Red Hat Enterprise Linux ti permette di gestire la piattaforma di IA come immagine di container, semplificando l'approccio alla scalabilità.

Accelerazione e strumenti IA

Per ottenere risultati in tempi più brevi, utilizza gli acceleratori hardware open source e le funzionalità ottimizzate di deep learning.

Scalabilità con Red Hat OpenShift AI

Red Hat OpenShift AI è una piattaforma MLOps integrata per la creazione, l'addestramento, l'ottimizzazione, la distribuzione e il monitoraggio su larga scala di applicazioni basate sull'IA e di modelli fondativi e predittivi in ambienti cloud ibridi.

Red Hat OpenShift AI si basa su Red Hat OpenShift per offrire un'esperienza coerente, semplificata e automatizzata per la gestione dei carichi di lavoro e dei requisiti prestazionali dei progetti di intelligenza artificiale/machine learning (AI/ML). Le pratiche MLOps possono aiutare le aziende a rispondere rapidamente alle innovazioni dell'IA e a produrre applicazioni basate su questa tecnologia più rapidamente.

Sperimenta l'ambiente sandbox di Red Hat OpenShift AI

Con i componenti ottimizzati di Open Data Hub e di altri progetti open source, Red Hat OpenShift AI offre a data scientist e sviluppatori un'efficiente piattaforma di AI/ML nel cloud ibrido, per creare applicazioni basate sull'IA partendo da dati fruibili.

[Provali nella sandbox per sviluppatori >](#)

Perché scegliere Red Hat OpenShift AI?

Distribuzione di modelli scalabili

I modelli possono essere distribuiti per l'integrazione in applicazioni intelligenti on premise, all'edge o nel cloud pubblico. Possono essere inoltre ricreati, ridistribuiti e monitorati in base alle modifiche del notebook sorgente.

Gestione semplificata dell'infrastruttura IA

Assicura ai tuoi team un accesso on demand alle risorse in modo che possano dedicarsi all'analisi dei dati e alla realizzazione di app in grado di offrire all'azienda un valore aggiunto effettivo.

Strumenti di AI/ML supportati e testati

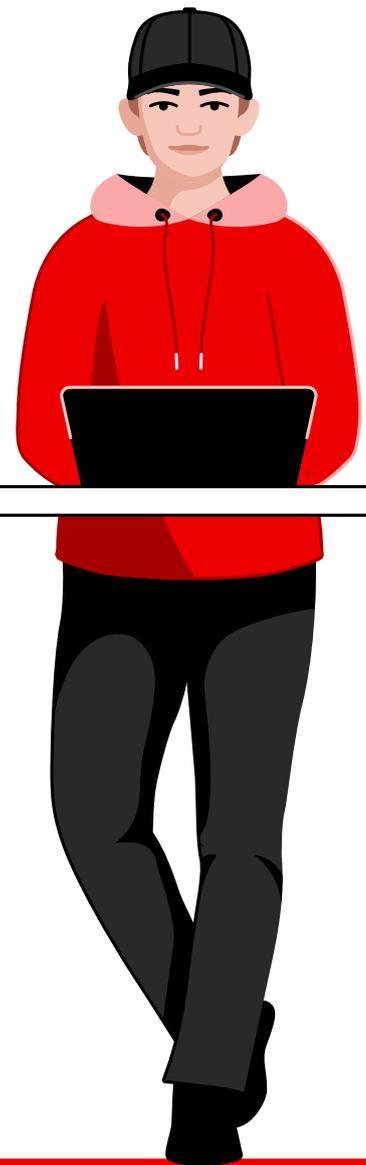
Lascia che sia Red Hat a monitorare, integrare, verificare e supportare i comuni strumenti e modelli di AI/ML sulla piattaforma applicativa Red Hat OpenShift.

Flessibilità nel cloud ibrido

Disponibile come software autogestito o come servizio cloud completamente gestito su Red Hat OpenShift, Red Hat OpenShift AI offre una piattaforma flessibile e sicura che ti permette di scegliere se sviluppare e distribuire i modelli on premise, nel cloud pubblico o all'edge.

Segui le nostre best practice

Red Hat Consulting offre servizi che consentono di installare, configurare e sfruttare al massimo Red Hat OpenShift AI, mettendo a disposizione supporto e indicazioni per qualsiasi tua iniziativa, dalla realizzazione di un progetto pilota con OpenShift AI alla creazione della tua base MLOps.



Red Hat offre una gamma completa di soluzioni tecnologiche, una comprovata esperienza e partnership strategiche per aiutarti a raggiungere gli obiettivi previsti. Fornisce una base per lo sviluppo e il deployment delle applicazioni e dei modelli di IA generativa, nonché servizi e percorsi formativi per accelerare l'adozione.

Scopri di più

Passa alla fase successiva del percorso di adozione dell'IA

Accelera l'adozione dell'IA con la strategia di Red Hat per il cloud ibrido, improntata alla flessibilità, con cui eseguire le applicazioni IA ovunque sia necessario.

Dai impulso ai tuoi progetti AI/ML con l'esperienza e i servizi di consulenza e formazione di Red Hat, che aiutano la tua azienda a raggiungere i risultati auspicati con l'IA.

Scopri di più sui servizi di AI/ML:

red.ht/aiml-consulting

Prenota una discovery session gratuita:

redhat.com/consulting

Scopri di più su Red Hat AI



[Scopri di più su Red Hat Enterprise Linux AI](#)



[Scopri la scalabilità con Red Hat OpenShift AI](#)