DELL Technologies NVIDIA

Sblocca risultati migliori

Dell Generative AI Solutions offre tutto ciò di cui hai bisogno per l'inferenza LLM (Large Language Model)

Genera previsioni e output con time-to-value più rapido, accelerando al contempo il processo decisionale con una soluzione GenAI di Dell Technologies e NVIDIA. Questa architettura congiunta offre una progettazione modulare, protetta e scalabile che supporta numerosi casi d'uso di inferenza e requisiti di elaborazione.

Ora puoi semplificare l'adozione della GenAI all'interno della tua organizzazione e ridurre il tempo necessario per ottenere i risultati con una soluzione testata.

Crea nuovo valore con un'infrastruttura protetta per le tue operazioni business-critical

Migliora il valore dell'IT con indicazioni strategiche

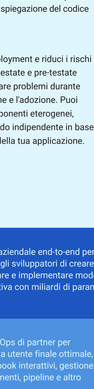
Dimensiona correttamente la tua infrastruttura con modelli di consumo flessibili

Cos'è l'inferenza?

In riferimento all'AI, l'inferenza indica il processo che prevede l'uso di un modello addestrato a generare previsioni, prendere decisioni o produrre output in base ai dati di input. Prevede l'applicazione delle conoscenze e dei pattern acquisiti durante la fase di addestramento del modello a dati nuovi e sconosciuti.

Durante l'inferenza, il modello addestrato acquisisce i dati di input e li elabora mediante i propri algoritmi computazionali o l'architettura della rete neurale per produrre un output o una previsione. Il modello applica le ponderazioni, le regole o i parametri appresi per trasformare i dati di input in informazioni o azioni significative.

L'inferenza è una fase cruciale nel ciclo di vita di un sistema AI. Dopo aver addestrato un modello su dati con o senza etichetta per apprendere i pattern e le correlazioni, l'inferenza permette al modello di generalizzare la propria conoscenza e di fare previsioni o generare risposte su dati reali o sconosciuti.



Come stai utilizzando la GenAI?

Generazione di linguaggio naturale:

Puoi utilizzare i modelli per le attività di generazione di testo quali la scrittura di documenti, la generazione di dialoghi, il riepilogo o la creazione di contenuto

Chatbot e assistenti virtuali:

Rendi più efficienti gli agent delle conversazioni, i chatbot e gli assistenti virtuali generando risposte in linguaggio naturale in base alle istruzioni o alle query degli utenti

Sviluppo di codice:

Ottieni assistenza nello sviluppo di codice con nuove funzionalità quali il completamento del codice, la possibilità di generare test delle unità o una funzione di chat per la spiegazione del codice

Dell Validated Design for Generative AI with NVIDIA - Inferenza

Accelera il deployment e riduci i rischi con soluzioni testate e pre-testate ideate per evitare problemi durante la progettazione e l'adozione. Puoi utilizzare componenti eterogenei, scalabili in modo indipendente in base alle esigenze della tua applicazione.

Framework di AI generativa

AI conversazionale: NVIDIA Nemo

Framework aziendale end-to-end per consentire agli sviluppatori di creare, personalizzare e implementare modelli di AI generativa con miliardi di parametri.

Piattaforme AI Ops e ML Ops

NVIDIA AI Enterprise

Software AI Ops di partner per un'esperienza utente finale ottimale, inclusi notebook interattivi, gestione degli esperimenti, pipeline e altro ancora.

Infrastruttura software

NVIDIA Base Command Manager Essentials

Livello di orchestration e programmazione per l'esecuzione di addestramento AI, inclusi processi su più nodi e dimensionamento dell'inferenza.

Gestione dell'infrastruttura

OpenManage Enterprise
OneFS | CloudIQ

Strumenti di gestione Dell familiari che offrono monitoraggio proattivo e analisi predittiva, semplificando le operazioni dell'infrastruttura.

Infrastruttura hardware

Server, storage, rete Dell | GPU NVIDIA

Elaborazione con acceleratori
Server Dell PowerEdge R760xa con GPU NVIDIA A100 o H100

Storage
Supporto per Dell PowerScale, ECS e ObjectScale

Rete
Dell PowerSwitch S5232F-ON o S5248F-ON

Fornisci risultati più rapidamente con il nostro aiuto

Gli esperti Dell Services possono fornirti assistenza in ogni fase del tuo percorso verso la GenAI:

Definizione della strategia

Crea la tua roadmap per raggiungere gli obiettivi di innovazione delle entità interessate IT e aziendali

Implementazione

Stabilisci la piattaforma, sfruttando le Dell Validated Design per implementare la GenAI con inferenza hardware e software

Adozione

Accelera il valore dei tuoi casi d'uso implementando un modello di inferenza pre-addestrato

Scalabilità

Gestisci il portafoglio di innovazione con esperti tecnici residenti e offerte di formazione per sviluppare le competenze GenAI del tuo team

Dell Technologies e NVIDIA

Dell Technologies e NVIDIA collaborano per abilitare e accelerare i carichi di lavoro di AI generativa, fornire hardware e software convalidati dal reparto tecnico per accelerare i carichi di lavoro AI, ML e DL in modo da soddisfare le esigenze del cliente in ogni tipologia di business e settore verticale. Con Dell Technologies e NVIDIA, puoi implementare soluzioni AI per accelerare la Digital Transformation tramite dati in tempo reale che miglioranо il processo decisionale chiave, con soluzioni ottimizzate per un time-to-value più veloce dalle tue iniziative AI.

Ulteriori informazioni sulle soluzioni Dell

Contatta un esperto Dell Technologies

Visualizza altre risorse

Partecipa alla conversazione con #PowerEdge @DellTech

https://www.dell.com/ai
https://www.dell.com/en-us/lp/contact-us
https://infohub.delltechnologies.com/