TD SYNNEX

MLOps per l'IA

Destinazione AI: caso d'uso

Introduzione

MLOps per l'intelligenza artificiale integra i principi DevOps con le esigenze specifiche dell'apprendimento automatico, semplificando l'intero ciclo di vita del modello, dalla sperimentazione e formazione all'implementazione, al monitoraggio e al miglioramento continuo. Questo approccio sfrutta piattaforme di orchestrazione dei container come Kubernetes, notebook con controllo delle versioni per gli esperimenti e pipeline automatizzate per gestire riproducibilità, scalabilità e governance. MLOps consente ai team di iterare in modo efficiente sui modelli, di trasferirli senza problemi in produzione e di monitorarne le prestazioni in tempo reale. Integrando test automatizzati, CI/CD per i modelli e una registrazione affidabile, le organizzazioni garantiscono che le loro iniziative di intelligenza artificiale siano affidabili e reattive ai mutevoli modelli di dati. Le soluzioni dei principali fornitori in questo spazio includono AWS SageMaker, Microsoft Azure Machine Learning, Google AI Platform, IBM Watsonx, Nvidia AI Enterprise Suite, Dell AI Factory e HPE Private Cloud AI, ognuno dei quali offre ambienti integrati per supportare l'intero spettro di flussi di lavoro di apprendimento automatico.

Punti dolenti

Lo sviluppo e la gestione di modelli di apprendimento automatico comportano numerose sfide, dal monitoraggio incoerente degli esperimenti e dai problemi di controllo delle versioni alle complesse procedure di distribuzione. I flussi di lavoro tradizionali spesso creano compartimenti stagni tra gli scienziati dei dati e i team operativi, causando ritardi nel passaggio dei modelli dalla ricerca alla produzione. Senza processi standardizzati, diventa difficile mantenere la riproducibilità e monitorare le prestazioni del modello nel tempo. L'assenza di meccanismi di monitoraggio e test automatizzati può comportare una riduzione dell'accuratezza del modello, tempi di inattività più lunghi e problemi di conformità. Inoltre, la scalabilità dei modelli di apprendimento automatico in vari ambienti, garantendo al contempo la sicurezza dei dati e l'utilizzo efficiente delle risorse, aggiunge ulteriore complessità, rallentando in ultima analisi l'innovazione e riducendo l'agilità operativa.

Vantaggi della soluzione

L'integrazione delle pratiche MLOps nelle iniziative di intelligenza artificiale offre vantaggi trasformativi. Le pipeline standardizzate e i flussi di lavoro automatizzati riducono i tempi di produzione colmando il divario tra sperimentazione e implementazione. L'integrazione e la distribuzione continue (CI/CD) dei modelli consentono un'iterazione rapida e aggiornamenti tempestivi, garantendo che i modelli rimangano accurati e pertinenti. La riproducibilità migliorata e il controllo delle versioni migliorano la collaborazione tra i team, mentre i test e il

monitoraggio automatizzati aiutano a identificare rapidamente anomalie o degrado delle prestazioni. Semplificando la gestione delle risorse e garantendo operazioni sicure e conformi, le organizzazioni raggiungono una maggiore efficienza operativa, riducono i tempi di inattività e mantengono la conformità alle normative. Nel complesso, MLOps consente ai team di scalare le soluzioni di intelligenza artificiale in modo affidabile, promuovendo l'innovazione e offrendo un valore aziendale costante.

KPI interessati

Gli indicatori chiave di prestazione che traggono vantaggio da MLOps per l'intelligenza artificiale includono:

Frequenza di distribuzione del modello: aumento della velocità di spostamento dei modelli dallo sviluppo alla produzione.

Time-to-Market: riduzione del tempo trascorso dallo sviluppo del modello alla distribuzione in tempo reale.

Precisione e deriva del modello: il monitoraggio continuo garantisce che i modelli mantengano elevate prestazioni nel tempo.

Efficienza operativa: miglioramento della produttività nelle pipeline di addestramento e distribuzione dei modelli, riducendo l'intervento manuale.

Tempi di inattività e tassi di incidenti: frequenza inferiore di incidenti di produzione grazie ai test automatizzati e al monitoraggio proattivo.

Implementazione del caso d'uso

L'implementazione di MLOps per l'intelligenza artificiale richiede un approccio completo e multifase che integri tecnologia, processi e allineamento culturale

Pianificazione e allineamento delle parti interessate

Definire gli obiettivi: articolare chiaramente obiettivi quali la riduzione dei cicli di distribuzione, il miglioramento dell'accuratezza del modello e il miglioramento della collaborazione tra i team di data science e IT. Associarli a KPI misurabili.

Crea un team interfunzionale: coinvolgi data scientist, ingegneri ML, specialisti DevOps, addetti alla sicurezza IT e stakeholder esecutivi per garantire che tutti gli aspetti, dalla sperimentazione alla produzione, vengano affrontati.

Valutazione di fornitori e strumenti: valuta piattaforme MLOps come AWS SageMaker, Microsoft Azure Machine Learning, Google AI Platform, IBM Watsonx, Nvidia AI Enterprise Suite, Dell AI Factory e HPE Private Cloud AI per determinare la soluzione più adatta alla tua infrastruttura esistente.

Allocazione delle risorse e infrastrutture

Integrazione e gestione dei dati: consolida set di dati, registri degli esperimenti e artefatti del modello in un repository centralizzato con un solido controllo delle versioni e gestione degli accessi.

Pianificazione del budget: allocare risorse per licenze software, aggiornamenti dell'infrastruttura cloud o on-premise, formazione e supporto continuo. Valuta il ROI confrontando tempi di implementazione ottimizzati e tempi di inattività ridotti con i costi di investimento.

Sicurezza e conformità: implementare misure di sicurezza efficaci, tra cui crittografia, controllo degli accessi basato sui ruoli e framework di conformità, per proteggere i dati sensibili e la proprietà intellettuale durante l'intero ciclo di vita del modello.

Progetto pilota e prova di concetto

Seleziona un modello o un flusso di lavoro di destinazione: identifica un modello non critico o un segmento specifico del tuo flusso di lavoro ML per pilotare il processo MLOps.

Configura strumenti e pipeline: imposta la piattaforma scelta, integrando notebook per esperimenti, orchestrazione di container (ad esempio, Kubernetes), test automatizzati e sistemi di monitoraggio.

Parametri di base: documentare i parametri attuali relativi alla frequenza di distribuzione, alle prestazioni del modello e ai tassi di incidenti per stabilire parametri di riferimento per il miglioramento.

Esecuzione e adozione

Onboarding e formazione: condurre una formazione completa per tutti i team coinvolti, sottolineando nuovi flussi di lavoro, strumenti e best practice nel controllo delle versioni, nell'implementazione e nel monitoraggio dei modelli.

Cicli di feedback iterativi: stabilire sessioni di revisione regolari per raccogliere feedback, perfezionare i flussi di lavoro e adattare le pipeline automatizzate in base ai dati sulle prestazioni reali.

Implementazione su vasta scala: estendere gradualmente le pratiche MLOps ad altri modelli e progetti, garantendo un'integrazione perfetta con i sistemi di produzione esistenti.

Monitoraggio e gestione del cambiamento

Monitoraggio continuo: distribuisci dashboard in tempo reale per monitorare KPI quali frequenza di distribuzione, accuratezza del modello e tassi di incidenti, garantendo un rapido rilevamento delle anomalie.

Gestione del rischio: sviluppare protocolli per affrontare potenziali problemi quali la deriva del modello o le vulnerabilità della sicurezza, garantendo solidi piani di emergenza e supervisione manuale ove necessario.

Scalabilità e integrazione: garantire che il framework MLOps si adatti alla crescita dell'organizzazione integrandosi con sistemi IT più ampi, pipeline di dati e infrastrutture di sicurezza.

Manutenzione e miglioramento continuo

Aggiornamenti di modelli e pipeline: aggiorna e riaddestra regolarmente i modelli con dati nuovi e perfeziona le pipeline di distribuzione per adattarle ai requisiti in evoluzione e alle minacce emergenti.

Documentazione: conservare registrazioni dettagliate dei flussi di lavoro, delle best practice e delle procedure di integrazione per supportare il miglioramento continuo e facilitare gli audit.

Coinvolgimento della comunità e dell'ecosistema: interagisci con forum di settore, comunità di fornitori e gruppi di ricerca per rimanere aggiornato sulle tendenze emergenti, sugli strumenti e sulle migliori pratiche in ambito MLOps.

Panoramica delle offerte dei fornitori

- AWS SageMaker: offre una suite integrata per la creazione, la formazione e la distribuzione di modelli di apprendimento automatico con solide funzionalità MLOps.
- Microsoft Azure Machine Learning: fornisce strumenti completi per il monitoraggio degli esperimenti, CI/CD automatizzato per i modelli e framework di distribuzione sicuri.
- Google AI Platform: sfrutta Kubernetes e l'orchestrazione ML avanzata per l'addestramento e la distribuzione di modelli scalabili.
- IBM Watsonx: offre una gestione del ciclo di vita del modello di intelligenza artificiale di livello aziendale e l'integrazione di funzionalità di intelligenza artificiale generativa all'interno di una piattaforma sicura e scalabile.
- Nvidia AI Enterprise Suite: offre un set completo di strumenti di intelligenza artificiale ottimizzati per l'addestramento, la distribuzione e l'inferenza accelerati dei modelli su varie infrastrutture.
- Dell AI Factory: fornisce soluzioni hardware e software integrate per semplificare le operazioni di apprendimento automatico e garantire un'implementazione efficiente e scalabile dei modelli.
- HPE Private Cloud AI: supporta soluzioni MLOps on-premise o ibride con funzionalità avanzate di sicurezza, conformità e scalabilità.

Conclusione:

MLOps per l'intelligenza artificiale è fondamentale per colmare il divario tra lo sviluppo di modelli innovativi e implementazioni di produzione affidabili e scalabili. Standardizzando i flussi di lavoro, automatizzando i test e l'implementazione dei modelli e garantendo un monitoraggio e una governance solidi, le organizzazioni possono accelerare le proprie iniziative di intelligenza artificiale mantenendo elevati standard di prestazioni e sicurezza. Un processo di implementazione strutturato, che comprende pianificazione strategica, allineamento delle risorse, test pilota e miglioramento continuo, garantisce che la transizione a MLOps produca miglioramenti misurabili in termini di efficienza e qualità del modello. Sfruttando soluzioni integrate di fornitori come AWS SageMaker, Microsoft.