

Obtenha resultados melhores



O Dell Generative AI Solutions oferece tudo o que você precisa para as inferências de grandes modelos de linguagem

Aumente a qualidade, faça previsões para reduzir o time-to-value e obtenha resultados, além de acelerar a tomada de decisões com uma avançada solução de IA generativa de pilha completa da Dell Technologies e da NVIDIA. Essa arquitetura conjunta oferece um design modular, seguro e escalável que dá suporte a uma série de casos de uso de inferência e requisitos de computação.

Agora, você pode usar uma solução comprovada para simplificar a adoção da IA generativa em toda a sua organização e reduzir o tempo necessário para obter resultados.



Agregue mais valor com uma infraestrutura segura para suas operações essenciais aos negócios



Aumente o valor da TI com orientações estratégicas



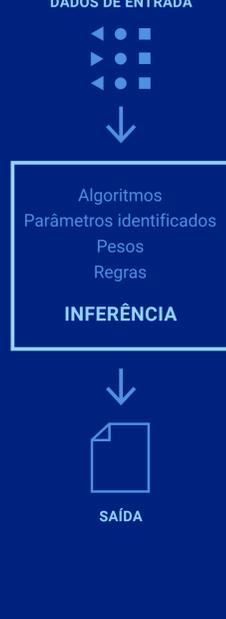
Dimensione a infraestrutura de maneira ideal usando modelos de consumo flexíveis

O que é inferência?

No âmbito da IA, inferência significa o processo de usar um modelo treinado para gerar previsões, tomar decisões ou produzir resultados com base em dados de entrada. Isso envolve aproveitar o aprendizado obtido e os padrões identificados durante a fase de treinamento dos modelos para aplicá-los a dados inéditos.

Durante a inferência, o modelo treinado é alimentado com dados e os processa por meio de algoritmos de computação ou de uma arquitetura de rede neural para produzir um resultado ou uma previsão. O modelo aplica os parâmetros, as regras ou os pesos identificados para transformar os dados de entrada em informações ou ações significativas.

A inferência é um estágio crucial do ciclo de vida dos sistemas de IA. Depois de treinar um modelo usando dados rotulados ou não rotulados para identificar padrões e correlações, a inferência permite que o modelo gere seu próprio conhecimento e faça previsões ou produza respostas para dados reais ou inéditos.



Como você está usando a IA generativa?

Geração de linguagem natural:

Os modelos podem ser usados em tarefas de geração de texto, como composição de documentos, geração de diálogos, produção de resumos ou criação de conteúdo

Chatbots e assistentes virtuais:

Potencialize agentes de conversa, chatbots e assistentes virtuais ao gerar respostas em linguagem natural com base em dúvidas ou instruções de usuários

Desenvolvimento de código:

Obtenha assistência no desenvolvimento de software usando recursos como conclusão de código, geração de testes unitários ou uma função de chat para explicar um código

Dell Validated Design for Generative AI with NVIDIA – inferência

Acelere as implementações e reduza os riscos com soluções pré-testadas e comprovadas, projetadas para ajudar a evitar armadilhas de design, planejamento e adoção. Os componentes podem ser combinados, assim como dimensionados de modo independente conforme as necessidades dos aplicativos.



Estrutura de IA generativa

IA conversacional: NVIDIA Nemo

Estrutura empresarial completa para que os desenvolvedores possam criar, personalizar e implementar modelos de IA generativa com bilhões de parâmetros.



Plataformas de AIOps e MLOps

NVIDIA AI Enterprise

Software de AIOps de parceiros para proporcionar uma experiência descomplicada aos usuários, inclusive notebooks interativos, gerenciamento de experimentos, pipelines e muito mais.



Infraestrutura de software

NVIDIA Base Command Manager Essentials

Camada de orquestração e programação para executar o treinamento de IA, inclusive trabalhos com vários nós e dimensionamento da inferência.



Gerenciamento de infraestrutura

OpenManage Enterprise OneFS | CloudIQ

Ferramentas de gerenciamento conhecidas da Dell para o monitoramento proativo e a análise preditiva, o que simplifica as operações da infraestrutura.



Infraestrutura de hardware

Dell Servers, Storage, Networking | NVIDIA GPUs

Computação com aceleradores
Servidores Dell PowerEdge R760xa com GPUs NVIDIA A100 ou H100

Armazenamento
Suporte para Dell PowerScale, ECS, e ObjectScale

Sistema de rede
Dell PowerSwitch S5232F-ON ou S5248F-ON

Conte com nossa ajuda para acelerar a obtenção de resultados

Os especialistas da Dell Services podem auxiliar em todos os estágios de sua jornada com a IA generativa:

Criar a estratégia

Crie seu roteiro para cumprir os objetivos de inovação das partes interessadas de TI e de negócios

Implementar

Estabeleça sua plataforma utilizando Dell Validated Designs para implementar soluções de IA generativa

Adotar

Amplie o valor de seus casos de uso ao implementar um modelo de inferência pré-treinado

Dimensionar

Gerencie seu portfólio de inovação aproveitando especialistas técnicos de treinamento para desenvolver habilidades de IA generativa em sua equipe

Dell Technologies e NVIDIA

A Dell Technologies e a NVIDIA trabalham em conjunto para habilitar e agilizar as cargas de trabalho de IA generativa, assim como oferecer hardware e software validados por engenharia para potencializar as cargas de trabalho de IA, ML e DL a fim de atender às necessidades dos clientes em todos os mercados verticais e empresas. Com a Dell Technologies e a NVIDIA, você pode implementar soluções de IA para acelerar a transformação digital por meio de dados em tempo real que aprimoram a tomada de decisões essenciais, além de soluções otimizadas para reduzir ao máximo o time-to-value das iniciativas de IA.



Saiba mais sobre as soluções Dell



Entre em contato com um especialista da Dell Technologies



Veja mais recursos



Participe da conversa com #PowerEdge e @DellTech