DELL Technologies

NVIDIA

Obtenez de meilleurs résultats

Dell Generative AI Solutions offre tout ce dont vous avez besoin pour l'inférence de grands modèles de langage

Générez des prédictions de délai de rentabilisation et des résultats plus rapides, et de meilleure qualité, tout en accélérant la prise de décision grâce à une puissante solution d'IA générative de Dell Technologies et NVIDIA sur la pile complète. Cette architecture conjointe offre une conception modulaire, sécurisée et évolutive qui prend en charge une multitude de cas d'utilisation d'inférence et d'exigences en matière de calcul.

Vous pouvez désormais simplifier l'adoption de l'IA générative dans votre organisation et réduire le temps nécessaire pour obtenir des résultats grâce à une solution éprouvée.

- Créez une nouvelle valeur avec une infrastructure sécurisée pour vos opérations stratégiques
- Améliorez la valeur IT avec des conseils stratégiques
- Dimensionnez correctement votre infrastructure avec des modèles de consommation flexibles

Qu'est-ce que l'inférence ?

Dans le domaine de l'IA, l'inférence désigne le processus consistant à utiliser un modèle entraîné pour générer des prédictions, prendre des décisions ou produire des résultats sur la base de données d'entrée. Cela consiste à appliquer à de nouvelles données inconnues les connaissances et les schémas acquis au cours de la phase d'entraînement du modèle.

Au cours de l'inférence, le modèle entraîné est alimenté par des données d'entrée, qu'il traite au moyen de ses algorithmes informatiques ou de son architecture de réseau neuronal pour produire un résultat ou une prédiction. Le modèle applique les paramètres qu'il a appris, des pondérations ou des règles pour transformer les données d'entrée en informations utiles ou actions.

L'inférence est une étape cruciale dans le cycle de vie d'un système d'IA. Après l'entraînement d'un modèle sur des données, étiquetées ou non, afin d'apprendre des schémas et des corrélations, l'inférence permet au modèle de généraliser ses connaissances et de faire des prédictions ou de générer des réponses sur des données du monde réel ou inconnues.

Comment utilisez-vous l'IA générative ?

Génération de langage naturel :
Les modèles peuvent être utilisés dans le cadre de tâches de génération de texte, par exemple la rédaction de documents ou de résumés, la génération de dialogues ou la création de contenu

Chatbots et assistants virtuels :
Optimise les agents conversationnels, chatbots et assistants virtuels grâce à la génération de réponses en langage naturel basées sur des requêtes ou des instructions de l'utilisateur

Développement de code :
Aide au développement logiciel avec des fonctionnalités comme la complétion de code, la possibilité de générer des tests unitaires ou une fonction de chat pour expliquer le code

Dell Validated Design for Generative AI with NVIDIA : inférence

Accélérez le déploiement et réduisez les risques avec des solutions prétestées et éprouvées conçues pour vous aider à éviter les pièges liés à la conception, à la planification et à l'adoption. Les composants peuvent être combinés et associés, et mis à l'échelle indépendamment selon vos besoins applicatifs.

Cadre de l'IA générative
IA conversationnelle : NVIDIA Nemo

Cadre d'entreprise de bout en bout permettant aux développeurs de créer, de personnaliser et de déployer des modèles d'IA générative avec des milliards de paramètres.

Plateformes AI Ops et ML Ops
NVIDIA AI Enterprise

Associez le logiciel AI Ops pour une expérience utilisateur final fluide : ordinateurs portables interactifs, facilité de gestion, développement d'opportunités, et bien plus encore.

Infrastructure logicielle
NVIDIA Base Command Manager Essentials

Couche d'orchestration et de planification pour entraîner l'IA, y compris sur les tâches à plusieurs nœuds, et la mise à l'échelle de l'inférence.

Gestion de l'infrastructure
OpenManage Enterprise
OneFS | CloudIQ

Outils de gestion Dell que vous connaissez déjà, permettant une surveillance proactive et une analytique prédictive qui simplifient les opérations d'infrastructure.

Infrastructure matérielle
Serveurs, stockage et gestion de réseau Dell | Processeurs graphiques NVIDIA

Calcul avec accélérateurs
Serveurs Dell PowerEdge R760xa avec des processeurs graphiques NVIDIA A100 ou H100

Stockage
Prise en charge de Dell PowerScale, ECS et ObjectScale

Gestion de réseau
Dell PowerSwitch S5232F-ON ou S5248F-ON

Fournissez des résultats plus rapidement avec notre aide

Les experts des services Dell peuvent vous aider à chaque étape de votre parcours d'adoption de l'IA générative :

Stratégie	Mise en œuvre	Adoption	Mise à l'échelle
Élaborez votre feuille de route pour atteindre les objectifs d'innovation de vos parties prenantes IT et commerciales	Établissez votre plateforme en utilisant les conceptions Dell Validated Designs pour mettre en œuvre le matériel et le logiciel d'inférence de l'IA générative	Accélérez la valeur de vos cas d'utilisation en implémentant un modèle d'inférence préentraîné	Gérez votre portefeuille de solutions d'innovation avec des experts techniques résidents et des offres de formation pour développer les compétences de votre équipe en matière d'IA générative

Dell Technologies et NVIDIA

Dell Technologies et NVIDIA travaillent ensemble pour prendre en charge et accélérer les charges applicatives d'IA générative, fournir du matériel et des logiciels validés par l'ingénierie pour accélérer les charges applicatives d'IA, de ML et de DL afin de répondre aux besoins des clients dans toutes les entreprises et tous les secteurs d'activité. Avec Dell Technologies et NVIDIA, vous pouvez déployer des solutions d'IA pour accélérer votre transformation numérique grâce à des données en temps réel qui améliorent la prise de décisions clés, avec des solutions optimisées pour accélérer le délai de rentabilisation de vos initiatives en matière d'IA.

En savoir plus sur les solutions Dell

Contacter un expert Dell Technologies

Afficher plus de ressources

Prenez part à la discussion avec #PowerEdge @DellTech

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

© 2023 Dell Inc. ou ses filiales. Tous droits réservés. Dell et les autres marques citées sont des marques commerciales de Dell Inc. ou de ses filiales. Les autres marques sont la propriété de leurs détenteurs respectifs.