Présentation de la solution



L’inférence d’IA à l’edge : moteur de la prochaine vague d’innovation

Une IA sécurisée et évolutive à l’edge grâce à HPE ProLiant Compute

Valeur commerciale de l'inférence d'IA à l'edge

Dans le monde centré sur les données d'aujourd'hui, la capacité à prendre des décisions en temps réel, là même où les données sont créées, est devenu un avantage déterminant. L'inférence d'IA à l'edge permet aux entreprises d'agir instantanément sur les informations issues des capteurs, caméras et devices connectés, offrant ainsi des niveaux jamais vus d'efficacité opérationnelle, d'expérience client et de résilience. En traitant les données localement, ces mêmes entreprises évitent les coûts et les délais liés à l'envoi de flux volumineux vers le cloud, tout en renforçant la confidentialité et la conformité.

L'essor de l'IA à l'edge s'accélère :

- Le marché mondial de l'IA à l'edge devrait atteindre 25 à 26 milliards de dollars en 2025, avec des taux de croissance annuels supérieurs à 20 % jusqu'en 2030.
- Plus de 90 % des DSI considèrent désormais l'IA à l'edge comme cruciale pour stimuler l'innovation et assurer la continuité de l'activité.
- Les cadres réglementaires (comme la loi européenne sur l'IA) exigent de plus en plus des inférences sur site auditables pour les charges de travail sensibles, ce qui rend les solutions edge non seulement stratégiques mais nécessaires.

L'inférence d'IA à l'edge est bien plus qu'une simple tendance technologique. C'est un catalyseur de croissance permettant aux entreprises de rivaliser, de s'adapter et de s'imposer dans un paysage en constante évolution.

Secteur	Principaux cas d'utilisation de l'edge et valeur ajoutée
Fabrication	**Contrôle qualité de la vision :** détection des anomalies en temps réel sur les lignes de production **Maintenance prédictive :** alertes précoces de défaillance d'équipement **Robotique autonome :** optimisation et sécurité des processus pilotées par l'IA
Défense	**Détection des menaces :** surveillance en temps réel et détection d'anomalies **Connaissance de la situation :** fusion de données pilotée par l'IA depuis plusieurs capteurs **Planification de la mission :** allocation automatisée des ressources et optimisation de la logistique
Télécommunications	**Optimisation du réseau :** allocation du trafic et des ressources gérée par l'IA **Détection de fraudes :** détection d'anomalies d'appels/de données en temps réel **Garantie de service :** prévention prédictive des interruptions
Vente au détail	**Prévention des pertes :** analyse vidéo instantanée pour la détection des vols **Analyse des rayons :** suivi automatisé des stocks **Caisse autonome :** expérience client fluide grâce à l'IA
Santé	**Analyse d'images :** diagnostic sur site pour un traitement plus rapide **Surveillance à distance :** alertes patient en temps réel **Suivi des actifs :** gestion efficace de la chaîne logistique et des équipements
Secteur public	**Gestion du trafic :** analyse vidéo de la fluidité et de la sécurité pour les villes intelligentes **Intervention d'urgence :** détection et envoi rapides des événements **Analyse des foules :** surveillance en temps réel de la sécurité publique
Énergie et services publics	**Surveillance du réseau :** détection des anomalies et optimisation de la performance réseau **État des actifs :** maintenance prédictive de l'infrastructure **Prévision de la demande :** équilibrage de charge et planification pilotés par l'IA
Transport et logistique	**Gestion de parc :** maintenance prédictive et optimisation des itinéraires **Suivi des marchandises :** surveillance en temps réel de la localisation et de l'état **Véhicules autonomes :** navigation et sécurité pilotées par l'IA
Hôtellerie-restauration	**Personnalisation de l'expérience client :** recommandations et services pilotés par l'IA **Sécurité et contrôle d'accès :** inférence vidéo en temps réel pour la sécurité **Optimisation des stocks :** gestion automatisée de la chaîne logistique
Enseignement	**Sécurité sur le campus :** analyse vidéo pour un accès sécurisé **Apprentissage adaptatif :** remarques en temps réel pour une formation personnalisée **Gestion des infrastructures physiques :** optimisation des ressources et de l'espace pilotée par l'IA

Prérequis techniques pour l'inférence à l'edge

L'efficacité de l'inférence d'IA à l'edge repose sur un ensemble d'exigences techniques fondamentales garantissant performance, fiabilité, sécurité et conformité dans les environnements distribués.

1. Performance et accélération

L'inférence à l'edge exige un traitement à haut débit et faible latence, capable de prendre en charge l'analyse et la prise de décision en temps réel. Les systèmes doivent gérer une accélération matérielle, comme celle des GPU ou des processeurs d'IA spécialisés, et permettre un calcul efficace pour les charges de travail telles que la vision par ordinateur et le traitement du langage naturel. Le support des types de données optimisés (par ex. INT8, FP16) et un refroidissement robuste sont essentiels pour une performance soutenue.

2. Évolutivité et flexibilité de déploiement

Les environnements edge sont très variés, allant des petits commerces de détail aux grandes infrastructures physiques industrielles en passant par les réseaux de télécommunications. L'infrastructure doit offrir des options modulaires et évolutives pour s'adapter à différentes tailles de déploiement et divers scénarios stratégiques, tout en supportant des configurations aussi compactes que robustes.

3. Sécurité et intégrité des données

Les fonctionnalités de sécurité intégrées au matériel comme la Silicon Root of Trust, le démarrage sécurisé et la validation du firmware d'exécution, contribuent à protéger les systèmes des falsifications et des accès non autorisés. Des mesures de sécurité physique peuvent également être prérequises pour prévenir le vol ou l'intrusion.

4. Gestion centralisée et automatisation

L'administration efficace des nœuds edge distribués nécessite des plateformes de gestion unifiée permettant une gouvernance centralisée du cycle de vie, une surveillance à distance, des mises à jour automatisées et l'intégration aux environnements de cloud hybride. Ces fonctionnalités permettent de réduire les frais généraux opérationnels et facilitent une résolution des incidents rapide.

5. Gouvernance et conformité

Les déploiements d'IA à l'edge doivent respecter des prérequis réglementaires en constante évolution, notamment en matière d'auditabilité et de confidentialité des données. Les solutions doivent prendre en charge l'orchestration fondée sur une politique, maintenir des pistes d'audit et permettre une gouvernance responsable de l'IA sur tous les sites.

6. Fiabilité et résilience opérationnelle

Un fonctionnement continu est essentiel, même en cas d'interruptions réseau ou de conditions défavorables. Pour limiter les temps d'arrêt et supporter les charges de travail stratégiques, l'infrastructure doit combiner conceptions robustes, options d'alimentation redondantes et fonctionnalités de gestion à distance.



Points clés à prendre en compte pour les DSI et praticiens de l'IA

Aligner l'IA à l'edge sur les résultats commerciaux Pour maximiser l'impact de l'IA à l'edge, les entreprises devraient privilégier les déploiements qui soutiennent directement leurs objectifs commerciaux majeurs comme la résilience opérationnelle, l'expérience client et la conformité réglementaire. Privilégiez les cas d'utilisation où la prise de décision en temps réel à l'edge produit des résultats mesurables (par ex. la maintenance prédictive, le contrôle de la qualité, l'analyse des données clients et l'optimisation réseau).

Bâtir en privilégiant l'évolutivité et la flexibilité Il est crucial de choisir une infrastructure edge capable de s'adapter à divers environnements tels que les usines, les points de vente, les bureaux distants et les réseaux de télécommunications. Votre solution doit pouvoir s'adapter à l'évolution des besoins de l'entreprise, aux nouveaux modèles d'IA et aux prérequis réglementaires changeants, garantissant ainsi sa pertinence et son agilité à long terme.

Simplifier la gestion et la gouvernance Il est crucial d'investir dans des plateformes offrant une gestion centralisée, une sécurité robuste et la conformité sur l'ensemble des sites edge. L'adoption de cadres de gouvernance solides contribuera à garantir un usage responsable de l'IA, à protéger la confidentialité des données et à maintenir la traçabilité (une nécessité face à l'évolution constante des réglementations comme la loi européenne sur l'IA).

Offrir aux équipes les moyens d'exploiter l'automatisation et les informations Tirez parti de l'automatisation et de l'observabilité pilotée par l'IA pour réduire les interventions manuelles, accélérer la résolution des problèmes et optimiser la performance. En réduisant les opérations de maintenance de routine et en permettant les opérations à distance, les équipes informatiques peuvent se concentrer sur des initiatives stratégiques qui stimulent la croissance commerciale.

Pérenniser vos investissements à l'edge Choisissez des partenaires et des solutions ayant fait leurs preuves en matière d'innovation de pointe, de sécurité et de support du cycle de vie. Envisagez des modèles de consommation flexibles et des services de conseil aptes à maximiser le ROI et l'adaptabilité de votre infrastructure à la croissance future et aux progrès technologiques.

Leçons de leadership

Bien plus qu'une simple mise à niveau technologique, l'inférence d'IA à l'edge est un levier stratégique pour la transformation de l'entreprise. En misant sur une infrastructure évolutive, une gestion centralisée, une gouvernance robuste et une innovation continue, les DSI et responsables IA peuvent créer de nouvelles sources de valeur, améliorer l'efficacité et positionner leurs entreprises pour un succès à long terme à l'edge.



Avantages de HPE ProLiant Compute

Sécurité multicouche

— La Silicon Root of Trust HPE iLO intégrée à tous les serveurs assure leur protection tout au long de leur cycle de vie, ainsi que leur préparation à la conformité en prévision des cyberattaques et attaques physiques futures

— Le démarrage sécurisé et la validation du firmware d'exécution empêchent toute falsification ou code non autorisé

— Des dispositifs de sécurité physique optionnels (kit anti-intrusion de châssis, verrou de façade, verrou Kensington) protègent les devices contre le vol ou l'intrusion

Performance et efficacité de pointe

— Puissance de traitement efficace pour gérer les charges de travail à faible latence, réduisant la dépendance à l'égard de l'infrastructure cloud

— Compatibilité IA avec support d'une large gamme de GPU NVIDIA® (L4, L40S, H100), pour alimenter l'inférence à l'edge dans les secteurs de la fabrication, du service public et de la défense

— Formats compacts et robustes pour un meilleur refroidissement, une meilleure circulation d'air et un fonctionnement silencieux ; capables de fonctionner dans des températures et des conditions extrêmes

Productivité et gestion opérationnelles

— HPE Compute Ops Management permet une gestion des opérations cloud robuste depuis une console unique, avec connexion sécurisée aux serveurs et gestion distribuée à l'échelle de l'entreprise

— La surveillance transparente des serveurs multifournisseur et les approbations de workflow réduisent les temps d'arrêt et maintiennent un fonctionnement optimal de l'infrastructure edge

— Réponse prioritaire avec accès à une assistance technique 24 h/24, 7 j/7 et 365 j/an

Conformité et chaîne logistique fiable

— Conformité à la loi TAA et chaîne logistique fiable avec des infrastructures physiques sécurisées de la fabrication à la livraison

— HPE offre une sécurité de serveur de pointe, de la puce au cloud, grâce à une sécurité avancée directement ancrée dans le silicium et à des ASIC sur mesure

Résultats quantifiables

— Risque réduit de falsification physique et numérique ; meilleure conformité et sécurité renforcée

— Réponse aux exigences de l'activité et alimentation efficace de toute charge de travail, n'importe où, pour aujourd'hui et pour l'avenir

— Gestion et contrôle simplifiés sur les environnements de serveurs distribués, offrant visibilité, informations et fonctionnalités automatisées pour une action immédiate



Façonner votre prochain kilomètre d'IA

L'inférence d'IA à l'edge remodèle la façon dont les entreprises opèrent, rivalisent et innovent en leur offrant une intelligence en temps réel, une résilience opérationnelle et une nouvelle valeur commerciale dès la création des données. Avec le portefeuille edge HPE ProLiant Compute spécialement conçu, une sécurité de pointe et une facilité de gestion éprouvée, vous êtes prêt à prendre un avantage décisif en cette nouvelle ère.

Que faire maintenant ?

- Évaluer vos opportunités à l'edge : identifiez les cas d'utilisation à fort impact où l'IA en temps réel peut générer des résultats mesurables dans votre entreprise
- Mobiliser vos parties prenantes : alignez les responsables IT, commerciaux et opérationnels autour d'une vision partagée, pour une transformation compatible avec l'edge
- Devenir partenaire pour plus de succès : tirez parti de l'expertise, des services et des modèles de consommation flexible de HPE pour accélérer vos déploiements et maximiser votre retour sur investissement
- Pérenniser vos investissements : choisissez une infrastructure évolutive, sécurisée et adaptable pour anticiper les exigences commerciales et réglementaires en constante évolution

Visiter HPE.com

Prêt à mettre l'IA à l'edge au travail ?

Explorez le portefeuille calcul edge computing HPE ProLiant Compute, accédez à des ressources sur les solutions et communiquez avec des experts.

Pour en savoir plus, rendez-vous sur

HPE.com/ProLiant/edge-computing

Live Chat

© Copyright 2026 Hewlett Packard Enterprise Development LP. Les informations figurant dans ce document sont susceptibles d'être modifiées sans préavis. Les seules garanties relatives aux produits et services Hewlett Packard Enterprise sont stipulées dans les déclarations de garantie expresses accompagnant ces produits et services. Aucune partie du présent document ne saurait être interprétée comme offrant une garantie supplémentaire. Hewlett Packard Enterprise décline toute responsabilité en cas d'erreurs ou d'omissions de nature technique ou rédactionnelle dans le présent document.

NVIDIA est une marque commerciale et/ou déposée de NVIDIA Corporation aux États-Unis et dans d'autres pays. Toutes les marques de tiers appartiennent à leurs propriétaires respectifs.

a00156356FRE

HEWLETT PACKARD ENTERPRISE

hpe.com