

Cloudflare for AI

Évoluez rapidement et maîtrisez les coûts tout en proposant des produits d'IA générative innovants

L'avenir des applications est intelligent

Développez localement, étendez à travers le monde

Les entreprises spécialistes de l'intelligence artificielle (IA) ont le vent en poupe, et un nombre croissant d'équipes de développeurs créent et étendent rapidement des applications d'IA générative. Pendant ce temps, les codeurs utilisent leurs propres capacités d'intelligence artificielle/d'apprentissage automatique pour améliorer leur productivité, contracter les courbes d'apprentissage et améliorer la qualité globale et la rapidité des projets de codage. Toutefois, ils sont confrontés à des défis importants et coûteux, tels que le développement de la capacité en GPU, le stockage d'immenses ensembles de données générées par les utilisateurs et l'apport de réponses aux nouvelles lacunes en matière de sécurité et d'exposition des données.

En élaborant des solutions sur le réseau de Cloudflare, les entreprises spécialistes de l'IA peuvent développer une architecture cloud rentable, qui évolue à leur rythme et dispose d'une capacité de calcul ultra-performante, soutenue par le réseau de Cloudflare. Les équipes peuvent réduire ou éliminer les frais de trafic sortant, tout en garantissant la sécurité et la fiabilité.



Lancez votre prochaine fusée avec Cloudflare

Cloudflare for Startups est notre initiative visant à aider les jeunes entreprises à accélérer et protéger leurs propriétés Internet. Notre objectif consiste à partager la puissance du réseau et des solutions Cloudflare avec la communauté des jeunes entreprises. L'accès à Cloudflare for Startups est disponible sur invitation uniquement. Les organisations intéressées peuvent postuler [ici](#).



IA serverless sur GPU

Avec Workers AI, exécutez des tâches d'inférence rapides, à faible latence, sur notre réseau mondial de GPU. Faites votre choix parmi les modèles les plus populaires, tels que Llama-2 et ResNet50. Les LLM prêts à l'emploi incluent la classification d'images et de textes, la recherche de similarités et la traduction.



Stockage et recherche d'intégrations

Accélérez et développez vos flux de travail IA avec Vectorize, notre base de travail IA avec Vectorize, notre base de données vectorielle. Stockez les intégrations nouvelles ou existantes afin de permettre la recherche sur vos données dans le cadre d'une utilisation répétée avec les modèles d'apprentissage automatique.



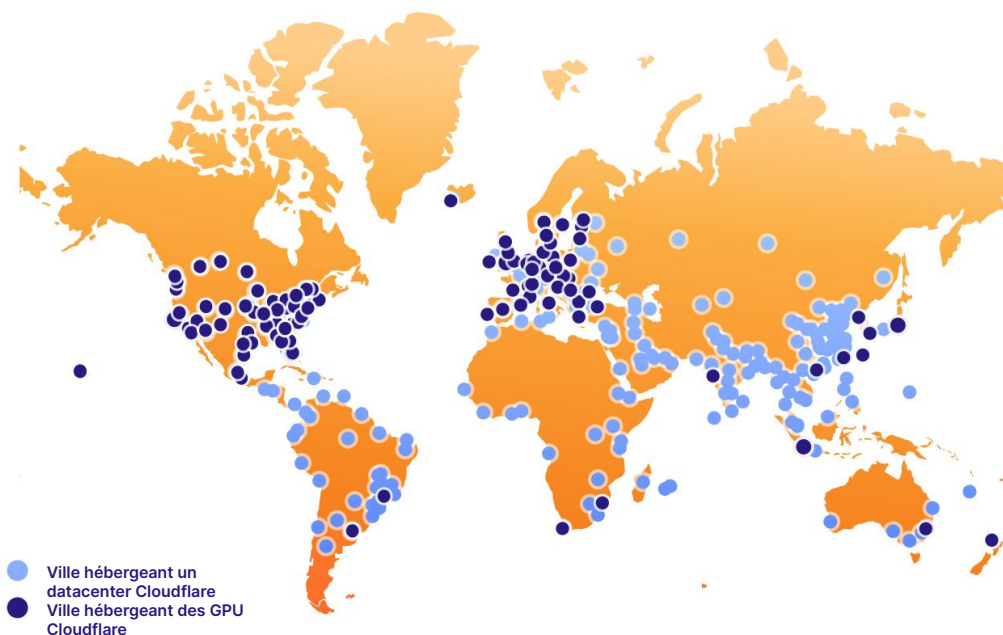
Contrôler et protéger

Protégez votre infrastructure back-end et mettez fin aux factures inattendues. Le service AI Gateway ajoute une couche de contrôle et de protection avec des limites de débit, la mise en cache et la visibilité du nombre d'utilisateurs des applications LLM.

Une infrastructure mondiale pour soutenir et sécuriser vos projets d'IA

Cloudflare est une solution de connectivité cloud conçue pour assurer la sécurité, la confidentialité, la rapidité et la fiabilité de toutes les ressources que vous connectez à Internet. Notre plateforme pour développeurs s'exécute sur notre réseau, fournissant une puissance de calcul à la périphérie et une capacité en GPU à l'endroit où vous en avez besoin. Vos équipes peuvent améliorer leur productivité et se dispenser des étapes inutilement chronophages en bénéficiant d'un chemin direct du code au déploiement serverless, avec une puissance de calcul et une capacité de stockage disponibles à la demande.

Sites de déploiement du réseau et des GPU



Fin 2023