

Cloudflare para IA

Escale rapidamente e gerencie custos enquanto fornece produtos inovadores de IA generativa

O futuro dos aplicativos é inteligente

Crie localmente e escale globalmente

As empresas de inteligência artificial (IA) estão crescendo e mais equipes de desenvolvedores estão criando e escalando rapidamente aplicativos de IA generativa. Enquanto isso, os programadores usam seus próprios recursos de IA/ML para aumentar a produtividade, reduzir as curvas de aprendizado e melhorar a qualidade geral e a velocidade dos projetos de codificação. Mas enfrentam desafios significativos e dispendiosos, como o dimensionamento da capacidade de GPU, o armazenamento de enormes conjuntos de dados gerados pelos usuários e a resolução de novas lacunas de segurança e exposição de dados.

Ao desenvolverem na rede da Cloudflare, as empresas de IA podem criar uma arquitetura em nuvem econômica que se adapta ao ritmo que precisam, com capacidade de computação de ponta líder do setor, alimentada pela rede da Cloudflare. As equipes podem reduzir ou eliminar custos de saída, garantindo segurança e confiabilidade.



Inicie seu próximo projeto de sucesso com a Cloudflare

O Cloudflare for Startups é nossa iniciativa para ajudar startups a acelerar e proteger seus ativos da internet. Nosso objetivo é compartilhar o poder da rede e das soluções da Cloudflare com a comunidade de startups. O Cloudflare for Startups é apenas por convite. As organizações interessadas podem se inscrever [aqui](#).



IA sem servidor em GPUs

Execute tarefas de inferência rápidas e de baixa latência em nossa rede global de GPUs com o Workers AI. Escolha entre modelos populares, incluindo Llama-2 e ResNet50. LLMs pré-criados incluem classificação, similaridade e tradução de imagens e textos.



Armazenar e pesquisar embeddings

Acelere e dimensione seus fluxos de trabalho de IA com o Vectorize, nosso banco de dados de vetores. Armazene embeddings novos ou existentes para permitir a pesquisa em seus próprios dados para uso repetido com modelos de aprendizado de máquina.



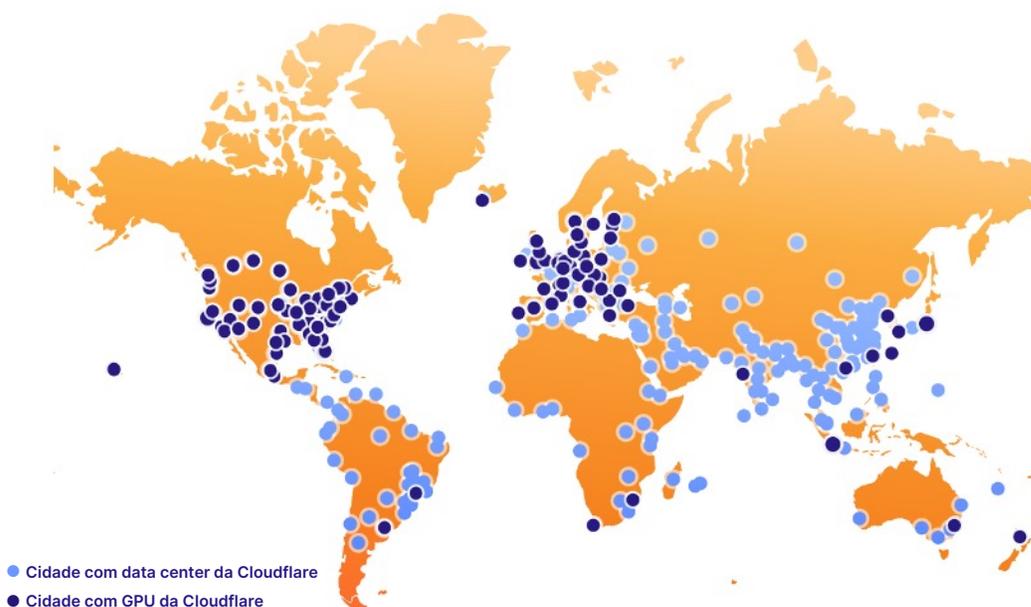
Controlar e proteger

Proteja a infraestrutura de back-end e evite contas inesperadas. O AI Gateway adiciona uma camada de controle e proteção com limites de taxa, armazenamento em cache e visibilidade de quantas pessoas estão usando aplicativos de LLM.

Infraestrutura global para potencializar e proteger seus projetos de IA

A Cloudflare é uma nuvem de conectividade projetada para tornar tudo o que você conecta à internet seguro, privado, rápido e confiável. Nossa plataforma para desenvolvedores é executada nessa rede para fornecer computação de ponta e capacidade de GPU sempre que você precisar. Suas equipes podem melhorar a produtividade e evitar etapas que desperdiçam tempo com um caminho direto do código até a implantação sem servidor, com computação e armazenamento disponíveis sob demanda.

Locais de rede e GPU



Final de 2023