
PRODUCT OVERVIEW

Runpod AI Developer Cloud overview

AI Developer Cloud for teams building, training, and deploying AI applications.

Pods

Dedicated GPU or CPU instances for containerized AI workloads.

Serverless

Pay-per-second compute with automatic scaling for production AI applications.

Public Endpoints

API access to pre-deployed AI models for image, video, audio, and text generation.

Why buyers use Runpod

Runpod gives AI teams a developer-first cloud for building, training, and deploying AI applications across pods, serverless workloads, public endpoints, and clusters.

Best-fit workloads

Training and fine-tuning, bursty inference, image and video generation, agent workloads, containerized AI applications, and distributed compute.

Buyer takeaway

For teams building production AI applications, Runpod is positioned as the AI Developer Cloud that helps developers keep compute flexible as workloads evolve.