

## PRODUCT OVERVIEW

# Runpod AI Developer Cloud overview

GPU infrastructure for developers building, training, and scaling AI applications.

### Pods

Dedicated GPU or CPU instances for containerized AI/ML workloads.

### Serverless

Pay-per-second compute with automatic scaling for production AI/ML apps.

### Public Endpoints

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

## Why buyers use Runpod

Runpod gives AI teams fast access to GPU infrastructure across pods, serverless workloads, public endpoints, and clusters, without forcing every team into long-term capacity planning before they know the shape of demand.

## Best-fit workloads

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

## Buyer takeaway

For teams that need GPU access without building their own infrastructure layer, Runpod is positioned as an AI Developer Cloud that helps developers ship faster while keeping compute flexible.