CAST AI

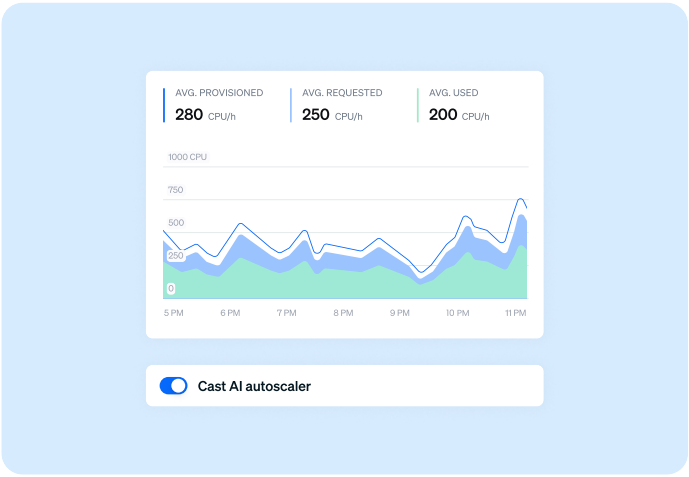
Intelligent Kubernetes automation on Google

Cast AI cuts cloud costs, secures Kubernetes clusters, and boosts DevOps efficiency.

4.8/5 50+ reviews

PLATFORM

Kubernetes automation platform

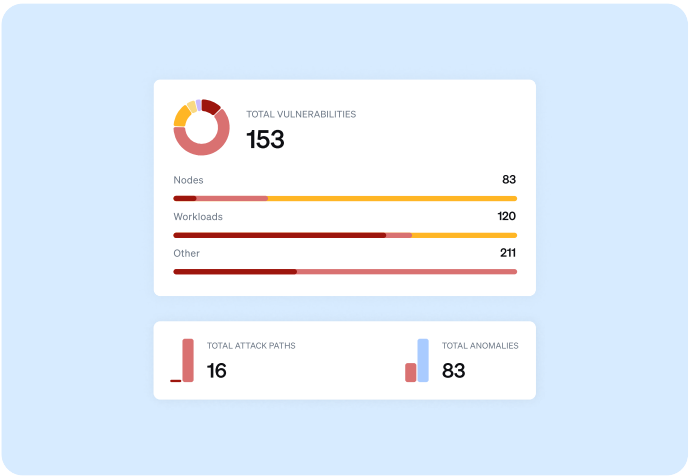


Automated Kubernetes cost optimization

Eliminate waste, maximize resource utilization, and scale resources up and down based on real-time requirements of your workloads.

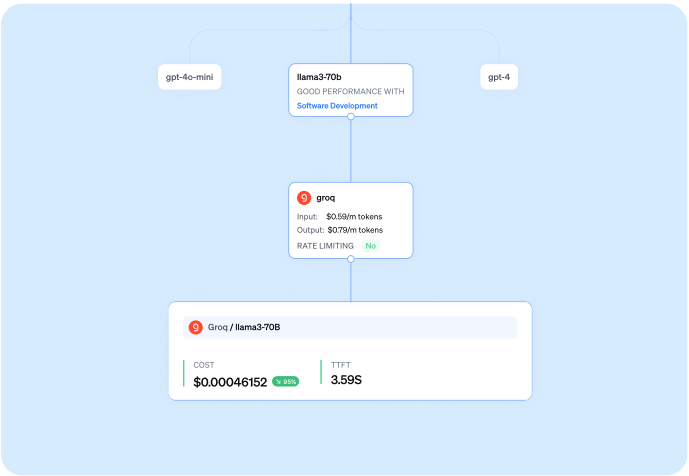
Kubernetes cost monitoring

Manage your Kubernetes expenses for free with detailed real-time tracking down to each namespace, workload, and resource allocation group.



Kubernetes and container security

Automatically detect, prioritize, and remediate critical issues to enhance your Kubernetes cluster's security posture, from scanning cluster images to identifying attack paths.



LLM optimization for AIOps

Automatically route user requests to the most optimal LLM for performance, cost and security

Cluster autoscaler

Eliminate waste and maximize resource utilization through bin packing, pod placement scheduling, and more.

Workload autoscaler

Automatically rightsize workload requests up or down to ensure optimal performance and cost-effectiveness.

Security posture

Continuously scan cluster images to identify vulnerabilities across each image layer.

Spot Instance automation

Automatically handle Spot Instance lifecycle events: interruptions, Spot diversity, and fallback to on-demand.

Cluster dashboard

Transform your Kubernetes management with real-time visibility into resource utilization across your clusters.

Cast AI Router

Automatically route your requests to the most optimal LLM, balancing performance, cost, and provider limits.

Available on:

Google Cloud Marketplace

Works with:

GCP

Integrates with:

Terraform, Helm, Grafana, Prometheus, Jira

Certified by:

ISO 27001, AICPA SOC 2

TESTIMONIALS

What our customers are saying:

Dekel Shavit
Senior Director of Engineering

Akamai

"The core savings we got are just brilliant, falling between 40-70%, depending on the workload. But that's not the full story. Before implementing CAST AI, my team was constantly moving around knobs and switches to make sure that our production environments and customers were up to par with the service we needed to invest in.

Now engineers have more bandwidth to focus on other areas they couldn't invest in before, like releasing features faster for our customers."

James O'Hare
Principal Platform Engineer

NIQ

"Cast AI paid for itself close to within the first two months thanks to all the savings it generated around compute costs."

Achi Solomon
Director of DevOps

yotpo.

"The ROI was great right from the start. Reducing 40% of our compute costs just by migrating our workloads to Cast AI—that's huge."

Boost Kubernetes performance, security, and cost optimization

Visit cast.ai to learn more