

## GPU capacity now, at a price your margins can carry.

GPCN gives AI software companies dedicated virtual GPU, CPU, and storage across 20+ countries, with one predictable price and a private deployment option for regulated customers.



### AT A GLANCE

**Challenge:** Scarce GPUs, costly inference, and demands for private deployment.

**GPCN answer:** Dedicated GPU capacity now, where customers require it.

**Result:** Better margins and faster entry into regulated accounts.

### BY THE NUMBERS

**53%**

Projected AI product gross margin in 2026, versus 75–85% for software.<sup>1</sup>

**23%**

Revenue spent on inference at scaling-stage AI companies.<sup>1</sup>

**49%**

Organizations naming GPU supply or power as their top compute barrier.<sup>2</sup>

### THE SITUATION

AI software companies live or die on compute. GPUs are rationed, hourly public cloud pricing moves with demand, and inference is a permanent cost of every customer interaction. At the same time, banks, hospitals, and governments now ask where the model runs before they buy. Growth depends on capacity you can count on and an answer regulated buyers accept.

### THE GPCN SOLUTION

GPCN is a network of private clouds giving AI companies simple access to dedicated virtual GPU, CPU, and storage across 60+ Tier 3+ data centers in 50+ cities and 20+ countries. Engineers spin up and manage their own virtual machines through a portal, API, or Terraform, with capacity available now.

- **One agreement:** every region, one relationship.
- **One price:** no egress fees within the network.
- **One SLA:** 99.99% availability, under one hour of downtime a year.
- **One access model:** portal, API, and Terraform.

**Typical fits:** training, inference, private customer deployments, and recovery sites.

### GPCN PERSPECTIVE

*"The AI companies that win regulated customers will be the ones that can answer "where does the model run?" with a dedicated, in-region answer."*

### WHO IT'S FOR

**CTO / Head of Engineering:** GPU capacity now, managed with Terraform.

**CFO:** predictable compute cost and healthier gross margin.

**Chief Revenue Officer:** a private deployment option for regulated buyers.

|                                   |                                          |                               |                                |
|-----------------------------------|------------------------------------------|-------------------------------|--------------------------------|
| <b>99.99%</b><br>Availability SLA | <b>SOC 2 Type II</b><br>Audited controls | <b>ISO 27001</b><br>Certified | <b>Tier 3+</b><br>Data centers |
|-----------------------------------|------------------------------------------|-------------------------------|--------------------------------|

### WHY GPCN IS DIFFERENT

|                             | Public Cloud                               | In-House / On-Prem                 | Other Private Cloud           | GPCN                                           |
|-----------------------------|--------------------------------------------|------------------------------------|-------------------------------|------------------------------------------------|
| <b>Cost model</b>           | Metered compute, storage, and egress       | Capital outlay, refresh, and staff | Custom quotes, opaque pricing | <b>One price, no egress fees in network</b>    |
| <b>GPU access</b>           | Quotas and waiting lists                   | 36–52 week lead times              | Limited footprint             | <b>Available now</b>                           |
| <b>Customer deployments</b> | Shared, multi-tenant                       | Your own hardware to maintain      | Varies by provider            | <b>Dedicated, in regions customers require</b> |
| <b>Availability</b>         | 99.5% standard: about 44 hours down a year | Your facility, your team           | Often one facility            | <b>99.99% SLA on Tier 3+ sites</b>             |
| <b>Operating model</b>      | Native-service lock-in                     | You run the hardware               | Vendor-specific tools         | <b>Portal, API, and Terraform</b>              |

Sources: (1) ICONIQ Capital, State of AI, January and July 2026. (2) Futurum Group, 2026. 99.99% availability equals about 53 minutes of downtime a year.

Available through EnTelegent Solutions | [gpcn.entelegent.com](https://gpcn.entelegent.com) | 800-975-7192 | 2026

Private by default. **Global by design.**