Cerebro
Cloud

The brain behind your cloud

This document has been prepared by Barrage d.o.o., the legal entity and owner of the Cerebro Cloud brand (together with its affiliates, the “Authors”), and is provided exclusively for the purpose of review. It is strictly private, confidential, and intended solely for the designated recipient(s). No part of this document may be copied, reproduced, stored, transmitted, or disclosed to any third party, in whole or in part, without the prior written consent of the Authors.

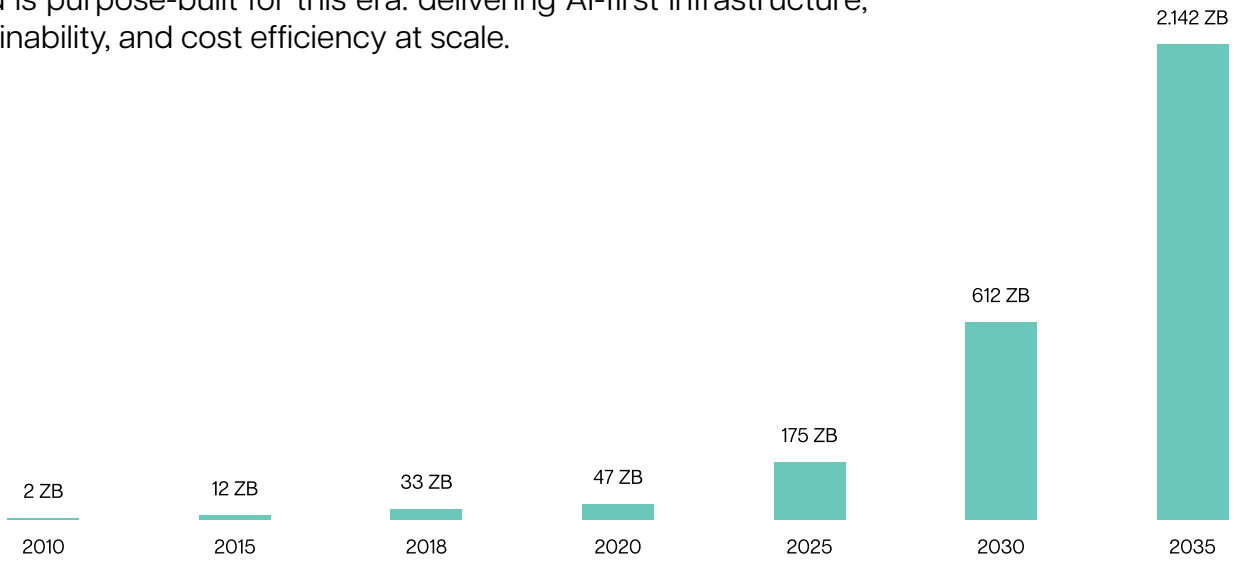
By accessing or reviewing this document, the recipient expressly acknowledges, accepts, and agrees to be bound by the confidentiality obligations set forth herein, as well as by any obligations arising under a valid Non-Disclosure Agreement (NDA) or applicable law.

Any unauthorized use, disclosure, publication, or distribution—whether direct or indirect—constitutes a material breach of confidentiality and will be deemed a serious violation of contractual and/or statutory obligations. Such breaches may result in monetary liability, injunctive relief, and enforcement under applicable commercial, civil, and obligations law.

Data deluge era

Unprecedented workloads demand powerful infrastructure.

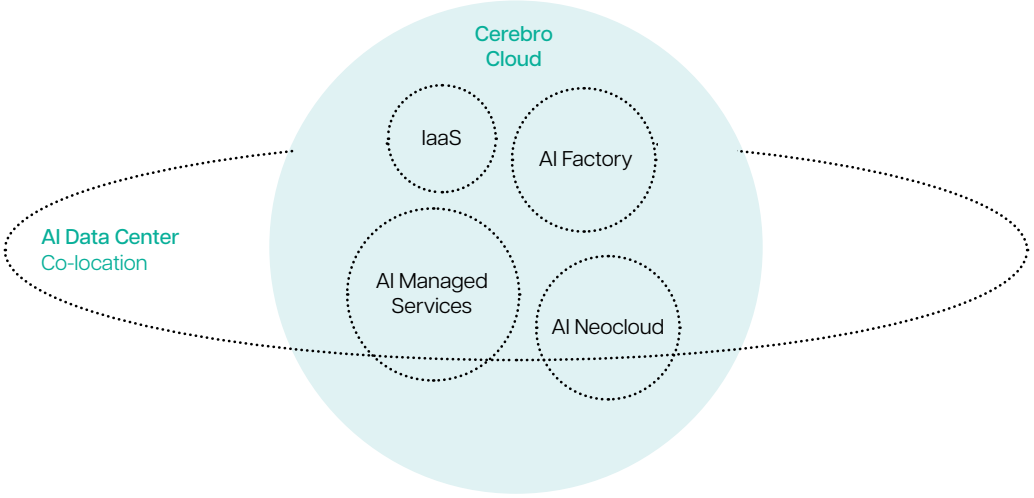
Global data creation is exploding – from 175 ZB in 2025 to 612 ZB by 2030. **AI workloads will account for 70% of this demand, requiring $6.7 trillion in new data center investment.** Cerebro Cloud is purpose-built for this era: delivering AI-first infrastructure, sustainability, and cost efficiency at scale.



What is Cerebro Cloud?

Cerebro Cloud is a full-stack AI platform designed for industrial-scale workloads. From silicon to solution, we provide high-density compute, managed orchestration, and AI-native services.

Launch bare-metal or virtual instances in minutes and focus on results – **we handle the complexity**.



Why choose Cerebro Cloud?

Sustainability
99% fossil-free power, PUE ≈ 1.1

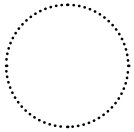
AI-optimized
200+ kW liquid-cooled racks, InfiniBand, NVIDIA Blackwell/Hopper

Simplicity
API-driven provisioning, real-time billing, multilingual support

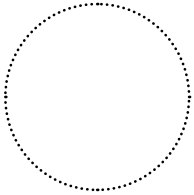
Scalability
Phased expansion to 200 MW of power

Data center location and capacity

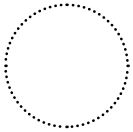
Hydrocompute 1
Sweden – 2 MW



Hydrocompute 2
Sweden – 4 MW



Hydrocompute 3
Sweden – 12 MW

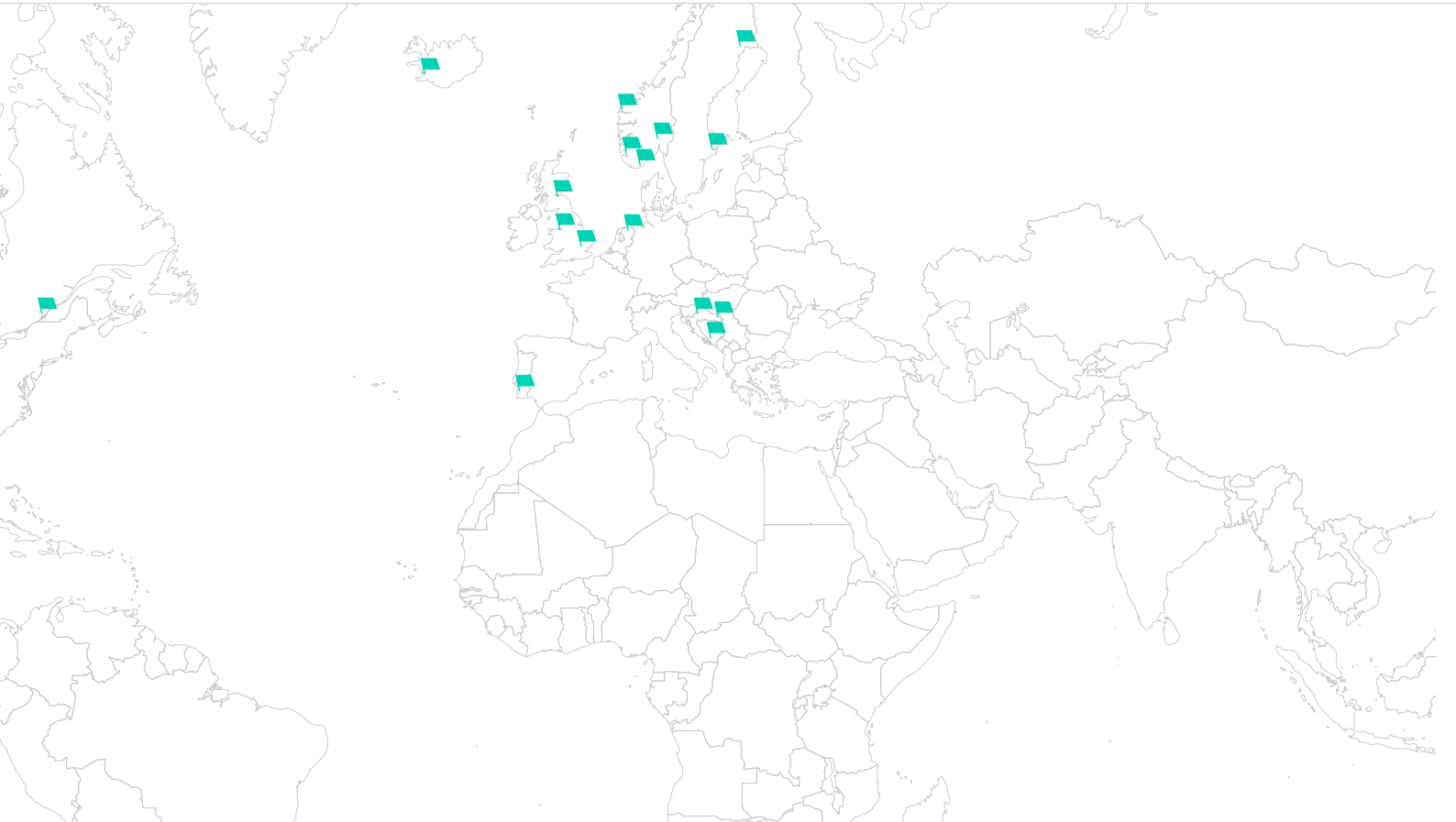


GridCompute
UK - 4 MW

Locations include:

- Tier III target with redundant A/B power feeds, advanced cooling, and sustainability initiatives, including heat recovery and a carbon-free energy footprint.
- Connectivity carrier-neutral facility with diverse fibre entry points and meet-me areas; certified to ISO27001 with programmes in place for PCI-DSS, SOC1/2, ISO14001, ISO50001, and ISO22301.
- CCTV and fire suppression; 24/7 staffed NOC, service desk, and remote hands.

Expansion pipeline: Nordics, Germany, the UK, the Netherlands, Portugal, Croatia.



Powering AI at every scale

From hardware to cloud services, our platform provides everything you need to build, run, and scale AI workloads:

Hardware	
High-performance GPUs	H100, H200, B200, A40, L40s, A6000
Next-gen platforms	GB200, GB300 NVL72
Ultra-dense compute	170+ kW racks, with advanced RDHx cooling

Software and services

Virtualization & orchestration	Proxmox, OpenStack, Kubernetes
Cerebro APIs	Provisioning, billing, orchestration
Cloud & AI services	Managed Kubernetes clusters, AI Factory, Neocloud APIs

Pricing

Resources available on demand:

Standard VM	*High-performance VM*	*High-end GPU VM*	*Premium bare-metal GPU*
$0.55/h	$1.50/h	$3.40/h	$6.60/h

Instances are provisioned and available for use in minutes.

Scaling is simple: get more GPU instances through a straightforward ordering process.

Batch AI discounts and custom configurations are available **upon request**.

Popular custom configuration deals:
Quantity: 256 GPUs, Term: 1 year, Storage: 30TB
Quantity: 64 GPUs, Term: month to month (min 3 months), Storage: 100TB

Cerebro Cloud - Easy-to-use customer portal

(available through invite only)

AI agents
Deploy with RAGU, connect APIs and data securely.

Creative workflows
GPU power for rendering, design, and simulations.

AI model inferencing
Fine-tune LLMs & vision models at scale.

Managed Kubernetes
Simplified orchestration for enterprise workloads.

Batch AI processing – save up to 50%

- Run prompts, generations, and enrichment jobs in scheduled windows
- Save ~50% with 24-hour processing
- Cross-cloud ready + fallback on-demand
- Cost control: budgets, alerts, usage breakdowns

Dedicated GPU clusters – scale seamlessly

- Provision scaling units (scale from 64 to 2048 or more GPUs) for critical workloads
- Flexible commitments (3 months to annual)
- Enterprise scale: hundreds of units
- Predictable pricing: fixed GPU-hour rates
- High availability: ready-to-deploy clusters

Spin up your AI infrastructure today.

Cerebro
Cloud

Tatjana Bašić
tatjana.basic@cerebro.cloud

Call / WhatsApp