GPU Day 2025

Massive parallel computing for science and industrial application



WAY TO THE CLOUD 99999.hu



Navigating the Al noise

with 99999 Informatika Kft.

Valentin Czuczumanov Head of Cloud and Al



Al innovations just shifted a gear!

- From monthly → weekly breakthroughs
- 58 notable AI models released in 2024 → 2025 is bringing a new class: reasoning-capable, agentic, and multimodal.
- These models aren't just chatting they're acting, planning, and integrating into workflows
- Huge investments and advancements in AI:

8. NVIDIA and the UAE secure U.S. approval to build a nice of Altraining cluster

The U.S. approved a deal that lets the UAE buy 500,000 NVIDIA chips per year. The chips will power a 5-gigawatt Al campus in Abu Dhabi, developed by G42 with support from U.S. firms like OpenAl



TRUMP SECURES \$600 BILLION SAUDI INVESTMENT COVERING AI, TECH, AND MORE

2. Google DeepMind unveils AlphaEvolve, a self-evolving codig agent that can find new algorithms

AlphaEvolve is designed to evolve codebases and discover new algorithms. It has already helped improve Google's data centers and chip designs, achieving up to 23 percent speed improvements

. MONT. ATTAL TIME, NAMES AND AND ADDRESS OF A				
ypers		Pagernine more services		
(and south that constant and configuration)				
"(where were, rere's a rere's a result of the second of the				
stantifi a second des des secondarias de secondarias				
r(seir) -= optas.oracientiranatoreation:				
all res into				
las(self.hypers.learning.rate)				
tend of the same tensor to the same set of the same set of the same set of	2010 C	2010		
dewidest Stypers.searning_rate, weight_decay-seit Stypers.weigh				
		design of the second second second	State and a second second	
salf) -> tax.no.initializers.foitializer:				
			and the second sec	
lalizer function.				
cars init scale				
he.e.mere				
pers normal(0 + 11 + 0, 1 + 11 + scale, inc. complexid)				
zers.normal(0 = 1) = 0, scale, jnp.complex64)				
		and the second se		
And the second s				
use(self, global_step, atart: float = 0.0, end: float = 0.0);				
al stan / salf config training stags			and the second s	
end) * frac + end				
		the second second second		
l(14x.111. static_aronums+0)				
	- Antonio antonio			
	1000			
: tuple inp.ndarray, inp.ndarray, inp.ndarray],		Terms		
And Continues				
tas.uptrate,				
ino.ndarray.				
ray,				
and the second se				
rray, jop.ndarray, jop.ndarray].				
*.				
p of decomposition parameter updates.				
and grassents.				
as unline and seally				
ar . retre "auto" "Auto"				
eposition, global_step, rng: 1np.mean(
and foldersectation whether start and				
uss_releasedeposition, grows.stap, reg)				
alabel ates and				
, gamma_step, fregi				
on real-valued functions of complex variables as must take				
int other control of control of the sector of the				
of the gradient.				
sufficient films and and an a state of the state and films a				
afore realized and a scould' dama				
des.				
the settle and the destruction and advect disconnected and				
ste = seif.opt.update(grade, opt_state, decomposition)				
ontax annly unfates(decompatition, undates)				
and a second sec				
the decomposition parameters.				

def _lineer_sd fran = 1 - g return (star Hunctools_part) de sult, de compositi ot_spata, fu de compositi ot_star global.step reg_;pp.ed tople[sp.ed ots.dylt] inp.ederrey. "-* tople! tople[sp.ed ots.dylt] sd.compte less loss, grab = linebad dec self.__

Iteration (

CONTROL OF ALL O

THAT THINK AS YOU GO



Mark Zuckerberg, CEO of Meta

Meet The MIT Dropouts Behind Cursor AI They built a \$2.6B coding powerhouse with barely 60 employees.

This team built custom AI models that make engineers 10X more productive.



e

replit



Agentic AI takes center stage

- From chatbots → autonomous agents
- Al that plans, decides & acts without waiting for prompts.
- 25 % of Gen-AI enterprises will pilot agentic AI in 2025

Deloitte 2025 TMT Predictions: "25 % of companies using generative AI will run agentic-AI pilots or PoCs in 2025."

• Many chooses already available:







FOMO says "GO!", worries shout "WAIT!

Everyone's racing into Al... but:

- Skills: thin talent pool
- Security: data & IP
- Availability: GPUs scarce
- Pace: week-to-week change
- Obsolescence: fear of sunk capex

99999 guides teams from AI hype to AI in production!





Who we are?





The cloud and AI Business Unit

TOur vision:

To empower Hungarian organizations to lead in digital transformation with local, secure, and scalable technologies.



Deliver cloud and AI solutions that are simple, practical, and aligned with real-world business goals.

💼 Our services:

• A **local cloud alternative** — secure, accessible, and built for Hungarian IT & compliance needs

 Practical Al engineering — design, build, and deploy enterprise-grade Al solutions using trusted partner tech (NVIDIA, Dell, HPE, IBM)





What is Skyblocks — and why another cloud?

Skyblocks is a cloud service designed and operated by us, created specifically for the Hungarian market.

🗹 Why we built it:

- A local alternative to hyperscalers with predictable pricing
- Designed for **data sovereignty**, faster adoption & **real support**
- Built on **well-known, enterprise-grade components** no black-box behavior
- Bridges the gap between **costly on-prem IT** and complex global clouds

🚀 What it offers:

- Reliable, modern infrastructure (VMware-based, ready-to-use)
- Expert team & local language support
- Transparent, OPEX-friendly pricing model
- Scalable compute, storage & networking on demand

Key Services of Skyblocks

Iaas (Infrastructure as a Service)

BaaS (Backup as a Service)

DRaaS (Disaster Recovery as a Service)

Cloud Security Services



Skyblocks: Core Services at a Glance

💻 laaS

• Enterprise IT at the push of a button

- VMware vSphere, vSAN, NSX, Cloud Director
- Built on HPE DL380 servers & Cisco networking
- Transparent pricing, 10 Gbps BIX connection

💾 BaaS

• Backup in the Cloud

- Self-service VM & file backup (on-site + cloud)
- Veeam integration, ransomware-proof secondary backup
- Scalable to 1 PB+, simple licensing

🔶 DRaaS

Disaster Recovery made flexible

- Veeam-based DR
- Self-service failover, flexible billing





Skyblocks: Security & Al Compute

Security from the Cloud

- SecOps & SOC as a Service (Fortinet + Andrews IT)
- Real-time monitoring, threat detection, NIS2-ready

🥥 GPUaaS

Power at your fingertips

- NVIDIA **L40s** available for training, fine-tuning, inference
- Simple provisioning, local support, low-latency access

• 🌠 Coming soon

- LLMaaS Managed access to tuned large models
- **RAGaaS** Secure retrieval-augmented generation service
- Agentic Al as a Service Autonomous systems you can deploy, evaluate, and control



Our Clients Speak for Us

Predictable Pricing

"We highly value the fact that costs are predictable upfront, and we don't have to worry about hidden charges during peak usage. That's why we are satisfied clients."



Access to Expertise

"What we value the most is that expert help is just a phone call away. This gives extra assurance that sets the service apart from others."

Tailored Solutions, Free POC

"Unlike most providers where we have to handle everything ourselves, here it is much simpler and more secure to get exactly what we need. We especially appreciated the option for a free POC (Proof of Concept) to test the service before making a decision."



Al at 99999 — From First Idea to **Real Impact**

What we offer today:

- Al consultancy identify real use-cases, not hype
- Infrastructure design & delivery tailored to business + data needs
- **Tech guidance** we bring the right tools from top enterprise vendors
- **Turnkey NVIDIA solutions** fast-start setups, fully supported

% What we're building next:

- **Ready-to-use AI solutions** complete stacks for common business problems
- Agentic Al systems next-gen automation, smart task execution
- Industry & edge deployments bringing AI to real-world environments



Real AI Projects We're Delivering Today

- Support Example 1 IT Operations Support
- Al-supported IT documentation updates
- Standardization & commenting of automation scripts
- Customer Example 2 Legal & Language Support
- Reviewing legal texts & contracts using GenAl
- Drafting structured **response letters** from input materials

🔢 Internal AI Use at 99999

- Designing Al-assisted work processes
- HR chatbot for internal Q&A
- Sales support platform using AI to assist our sales team



My experience from NVIDIA GTC2025





My Key Takeaways from GTC 2025

Reasoning > Just Text Generation

- Next-gen models don't just complete text they **think through steps**
- Token count is increasing as models reason in more steps and compute per token is also rising → inference is getting heavier on both fronts.

• 🝥 NVIDIA is also a software powerhouse

- Not just GPUs: full stack from NeMo to NIM, Riva, TensorRT, and the new Dynamo (formerly Triton)
- Developers can move from PoC to prod faster on any scale

The AI data center is evolving

- From "rack rooms" to **AI factories**
- AI-scale workloads now demand intensive power, liquid cooling, and smarter system management – (Digital twins in NVIDIA Omniverse)



DGX Spark: Local AI powerhouse for everyone

What it is:

A **palm-sized AI powerhouse** running up to **200B parameter models** locally.

🔋 Key specs:

- GB10 Grace Blackwell Superchip
- 1000 AI TOPS of performance
- 128 GB memory
- Compact, energy-efficient design



- Full **AI dev workflow** on-device: prototype → fine-tune → inference
- Supports reasoning-scale LLMs from Meta, Google, DeepSeek, and more
- Seamless hand-off to data center or cloud for scale-up





NVIDIA Quantum day

"This was the first event where a CEO invites everyone to explain why he was wrong."

Huang's shift:

Earlier: **20+ years** until useful quantum Now: "I underestimated it." Quantum is closer than we think

NVIDIA's response:

Announced the **NVAQC Research Center** (Boston) Built around **DGX Quantum + GB200 NVL72** architecture Bridges **quantum hardware** and **GPU compute** via **CUDA-Q** Focus: quantum error correction, hybrid apps, real-world use

💛 Why it matters:

Quantum startups + AWS + Microsoft all present **GPUs remain central** to simulating and scaling quantum workflows

NVIDIA positions itself as a **critical integration layer** in the future of computing.





So, how we help?

1. AI consultation

- Let's identify the most valuable use-cases together
- We assess feasibility before you invest

2. Cloud or not?

- Prefer public? → We set it up on **NVIDIA GPU Cloud** or **IBM Cloud**
- Prefer local? → We're offering our own AlaaS on Skyblocks with NVIDIA GPUs

3. On-prem deployment?

• We build with **trusted enterprise partners**:











Let's bring your AI ideas to life

Have an <mark>idea</mark>?

We'll help validate and shape it

Need infrastructure?

We'll design and deliver it



Want real outcomes?

We'll build with you — not just for you





Thank you for your attention!

Feel free to reach out to us with any questions: valentin.czuczumanov@99999.hu



