

Newsletter AI : Do Data Center à Secretária – A Revolução Silenciosa da IA

Publicado em 2025-10-20 10:02:27



Softelabs TechPulse – Outubro 2025

“Onde o silício encontra a inteligência”

1. NVIDIA redefine o conceito de workstation de IA

DGX Spark — o supercomputador de secretária com chip **Grace Blackwell GB10**, 128 GB de memória unificada e 1 petaFLOP FP4. Ideal para inferência, fine-tuning e agentes locais. Parcerias confirmadas: Dell, HP, ASUS e MSI.

DGX Station (GB300) — com 784 GB de memória unificada e performance multi-PFLOP, destina-se a equipas de IA e laboratórios. Marca a chegada da **IA híbrida** (local + cloud).

Tendência: democratização da IA de alto desempenho — o data center desce à secretária.

2. DeepSeek avança com eficiência e independência

- **DeepSeek-V3.2-Exp** compatível com chips Huawei Ascend e Cambricon.
- **DeepSeek-R1-0528** melhora raciocínio e modo thinking / non-thinking.
- **DeepSeek-R2** atrasado por limitações do hardware nacional.
- Custo de treino do R1: **~US\$ 294 000**.
- APIs com descontos off-peak até **–75 %**.

Tendência: modelos de raciocínio cada vez mais eficientes e acessíveis — o bloco de Leste da IA ganha força.

3. Competição: o tabuleiro global aquece

Ecosistema	Força dominante	Fragilidade
NVIDIA + OpenAI / Anthropic	Estabilidade, software maduro	Custo alto, dependência cloud

DeepSeek / Huawei / Cambricon	Eficiência, custo/ compute	Segurança fraca, maturidade limitada
Meta (Llama- Next)	Open-source, comunidade	Menor precisão factual
Google Gemini 2.5 Pro	Multimodalidade de topo	Custo operacional elevado

4. Impacto para PME e equipas técnicas

Hardware local (**DGX Spark / DGX Station**) permite **treinar e servir modelos com dados próprios**. Modelos como **DeepSeek-V3.2** ou **Llama-3** cabem em servidores de 128 GB GPU com quantização 4-8 bits. A solução híbrida (local + cloud) torna-se padrão — equilibrando custo, privacidade e latência.

5. Perspectiva Estratégica 2026

1. Modelos **open-weights** tornam-se padrão empresarial.
2. Estações locais substituem parte dos clusters cloud.
3. DeepSeek R2 / V4 rivalizam GPT-6 e Claude 4.
4. Chegada do chip **NVIDIA Rubin** com +45 % de eficiência FP8.
5. Europa e China investem em **soberania tecnológica**.

“O poder da IA está a sair dos data centers e a pousar nas secretárias — e quem souber usá-lo com ética e

engenho terá a vantagem.”

— *Augustus Veritas, Outubro 2025*

© 2025 Softelabs • Investigação e Desenvolvimento em IA
www.softelabs.pt • infos@softelabs.pt



Fragmentos do Caos:

[Blogue](#)

•

[Ebooks](#)

•

[Carrossel](#)



Esta página foi visitada ... vezes.

[Contactos](#)