

Blogue Fragmentos do Caos



A verdade nasce onde o pensamento é livre.

O Ferro que Pensa: DGX Spark/Station e DeepSeek — a IA entra na oficina do mundo

Publicado em 2026-01-26 14:13:53



BOX DE FACTOS

- **DGX Spark:** “personal AI supercomputer” com **GB10 Grace Blackwell**, **128GB** de memória unificada coerente e referência de preço **\$3,999** no marketplace NVIDIA. ^{o~}
- **DGX Spark:** especificações e mensagens de mercado repetem o alvo: **até ~200B parâmetros** em

Blogue Fragmentos do Caos



A verdade nasce onde o pensamento é livre.

Ultra, até **20 PFLOPS** e **784GB** de memória

unificada; networking até **800Gb/s** (ConnectX-8).²

- **DeepSeek-V3.2** (1 Dez 2025): integra “thinking” directamente em **tool-use** e suporta ferramentas em modos “thinking” e “non-thinking”.³

- **DeepSeek-V3.2-Exp** (29 Set 2025): passo “intermédio” com **DeepSeek Sparse Attention (DSA)** para contexto longo e corte de preço de API >50%.⁴

- **DeepSeek V4**: Reuters reporta que poderá sair em **meados de Fevereiro de 2026**, com foco em **coding** e prompts longos (a DeepSeek não confirmou oficialmente).⁵

O Ferro que Pensa: DGX Spark/Station e DeepSeek — a IA entra na oficina do mundo

A IA esteve anos a viver no céu das APIs — distante, cara, e frequentemente opaca. Agora desce à secretária,

Blogue Fragmentos do Caos



A verdade nasce onde o pensamento é livre.

Imagem — IA 2026 (substituir pelo URL depois de carregares no WordPress).



1) DGX Spark: o “supercomputador pessoal” deixa de ser metáfora

A NVIDIA pôs nome ao que muitos andavam a improvisar: a ideia de que o programador, a PME e o laboratório não podem viver eternamente dependentes do taxímetro da cloud. O **DGX Spark** surge como peça de ruptura — um “personal AI supercomputer” com referência de preço **\$3,999** no marketplace da NVIDIA, e uma promessa que tem algo de subversivo: **potência local, memória unificada** e um ecossistema que não pede licença para existir.⁶

Blogue Fragmentos do Caos



A verdade nasce onde o pensamento é livre.

desenvolvimento e teste (dependendo de quantização e arquitectura do modelo).⁷ Ou seja: não é um brinquedo. É um nó de IA que começa a parecer infra-estrutura.

2) DGX Station: quando a secretária começa a cheirar a data center

Se o Spark é a democratização, a **DGX Station** é a afirmação: o desktop como habitat natural para treino e inferência sérios. A NVIDIA descreve a Station como o primeiro sistema construído com o **GB300 Grace Blackwell Ultra**, oferecendo até **20 PFLOPS** e **784GB** de memória unificada — com **ConnectX-8** a abrir portas a escalabilidade multi-station via **800Gb/s**.⁸

Há aqui uma mensagem que interessa ao mundo real: a NVIDIA está a dizer que “IA grande” já não pertence apenas aos gigantes. Pertence também a equipas pequenas — desde que tenham disciplina, arquitectura e coragem para trazer o poder para dentro de casa.

Blogue Fragmentos do Caos



A verdade nasce onde o pensamento é livre.

promete e pouca gente entrega: aproximou o “raciocínio” do “fazer”. O **DeepSeek-V3.2** afirma-se como o primeiro modelo da casa a integrar “thinking” directamente em **tool-use**, e a suportar ferramentas tanto no modo “thinking” como no “non-thinking”.⁹

Traduzido para o chão do negócio: isto é a ponte entre **responder** e **operar**. Entre escrever bonito e produzir trabalho: chamar APIs internas, consultar bases de dados, gerar propostas, validar regras, compor relatórios — com guardrails, claro.

4) DSA e o contexto longo: quando a memória deixa de ser um luxo

O **V3.2-Exp** foi apresentado como passo “intermédio”, mas trouxe uma peça-chave: **DeepSeek Sparse Attention (DSA)**, desenhada para reduzir custo computacional em texto longo e pressionar preços (incluindo um corte de API >50% anunciado pela empresa).¹⁰

E é aqui que Spark/Station e DeepSeek se olham nos olhos: modelos que respiram melhor em contexto longo + hardware com memória coerente generosa = terreno fértil para **RAG empresarial** (documentação, procedimentos,

Blogue Fragmentos do Caos



A verdade nasce onde o pensamento é livre.

assinatura (mas ainda sem carimbo oficial)

Há rumores e há rumores com peso. A Reuters reportou (com base num relatório do The Information) que a DeepSeek poderá lançar o **V4** em **meados de Fevereiro de 2026**, com forte foco em **coding** e capacidade para processar prompts de código muito longos — algo que pode mexer com produtividade real em engenharia de software. A DeepSeek não confirmou oficialmente, por isso a data deve ser tratada como provisória, mas a direcção é clara.^{11~}

Epílogo: 2026 é o ano em que a IA volta a ser infra-estrutura

Há uma mudança silenciosa a acontecer: a IA deixa de ser um “serviço distante” e passa a ser uma peça do teu ecossistema — ao lado do teu servidor, do teu NAS, do teu ERP, do teu código. O **DGX Spark** abre a porta. A **DGX Station** arromba o portão.^{12~}

E do lado dos modelos, a DeepSeek insiste na mesma heresia produtiva: não basta falar, é preciso fazer — com eficiência, contexto longo e uso de ferramentas. E é nessa combinação — ferro e raciocínio — que 2026 começa a ganhar forma.



Blogue Fragmentos do Caos

A verdade nasce onde o pensamento é livre.

[leia]



Fragmentos do Caos:

[Blogue](#)

•

[Ebooks](#)

•

[Carrossel](#)



Esta página foi visitada ... vezes.

[Contactos](#)