

DeepSeek — Modelli API per Nexus OS™

Consigliato per Nexus: `deepseek-v4-pro`

Il rapporto qualità/prezzo più estremo dell'intero panorama AI. A \$0.435/\$0.87 costa 17 volte meno di GPT-5.6 Terra sull'output e 11 volte meno di Sonnet 5, con una qualità che regge il confronto nella Round Table (80.6% su SWE-bench Verified). Context window da 1M token, output fino a 384K, open weights (MIT). Nella peer review di Nexus, DeepSeek V4 Pro è costantemente tra le risposte migliori a una frazione del costo degli altri provider.

Nota: L'alias `deepseek-chat` è stato deprecato il 24/07/2026. Usare esclusivamente `deepseek-v4-pro` o `deepseek-v4-flash`.

Modelli disponibili via API — Luglio 2026

Prezzi Standard tier, per 1M token, in USD. Context 1M flat. Endpoint OpenAI-compatible: `api.deepseek.com`

Stringa API	Input	Output	Context	Forza principale	Note
<code>deepseek-v4-pro</code> ★	\$0.435	\$0.87	1M	Flagship: coding, reasoning, qualità near-frontier	Consigliato — Open weights (MIT). Max output 384K. Cache hit: \$0.004/1M input
<code>deepseek-v4-flash</code>	\$0.14	\$0.28	1M	Volume alto, task semplici e classificazione	Max output 384K. Cache hit: \$0.003/1M input. 2,500 req concorrenti

Dati verificati sulla documentazione ufficiale DeepSeek — `api-docs.deepseek.com` — Luglio 2026