LLM Verfügbarkeitsübersicht

aihub.biteno.com | Stand: 2026-11-03 | 20 lokal + 75 remote Modelle

Lokal gehostete Modelle (20)

Modell	Input / 1M Token	Output / 1M Token	Kontext
bge-m3	$0.2500	$0.5000	—
deepseek-r1:14b	$0.5000	$0.9000	—
deepseek-r1:32b	$0.5000	$0.9000	—
gemma3:27b	$0.5000	$0.9000	—
gemma4:31b	$0.3000	$0.9000	—
glm-4.7-flash:latest	$0.2500	$0.5000	—
granite-4.1-8b	$0.2500	$0.4000	—
laguna-xs-2.1	$0.2000	$0.4000	—
lfm2.5-thinking:1.2b	$0.5000	$0.9000	—
lfm2:24b	$0.5000	$0.9000	—
llama3.1-8b	$0.2500	$0.4000	2k
minimax-m2.7	$0.3500	$1.2500	—
mistral-small-24k:24b	$0.5000	$0.9000	—
mistral-small3.2-32k:24b	$0.2500	$0.5000	98k
nemotron-3-nano:4b	$0.2500	$0.4000	—
phi4:latest	$0.2500	$0.5000	—
qwen3-vl-8B-instruct	$0.2500	$0.4000	—
qwen3.6:35b-a3b	$0.5000	$0.9000	—
qwen3:14b	$0.2500	$0.5000	—
qwen3:32b	$0.5000	$0.9000	—

Remote / API Modelle (75)

Modell	Input / 1M Token	Output / 1M Token	Kontext
remote-claude-3-5-haiku	—	—	—
remote-claude-3-5-sonnet	—	—	—
remote-claude-3-7-sonnet-latest	—	—	—
remote-claude-4-sonnet-20250514	—	—	—
remote-claude-fable-5	$10.0000	$50.0000	128k
remote-claude-fable-5.1	$10.0000	$50.0000	1M
remote-claude-haiku-4-5	$1.0000	$5.0000	200k
remote-claude-haiku-5-5	$0.5000	$2.5000	128k
remote-claude-haiku-5-5-xhigh	$0.5000	$2.5000	128k
remote-claude-opus-4-5	$5.0000	$25.0000	200k
remote-claude-opus-4-6	$5.0000	$25.0000	1M
remote-claude-opus-4-7	$5.0000	$25.0000	1M
remote-claude-opus-4-8	$5.0000	$25.0000	128k
remote-claude-opus-5	$5.0000	$25.0000	200k
remote-claude-opus-5-5	$4.0000	$20.0000	128k
remote-claude-sonnet-4-5	$3.0000	$15.0000	200k
remote-claude-sonnet-4-6	$3.0000	$15.0000	1M
remote-claude-sonnet-5	$2.0000	$10.0000	64k
remote-claude-sonnet-5-5	$2.0000	$10.0000	128k
remote-deepseek-chat	$0.2800	$0.4200	131k
remote-deepseek-coder	$0.1400	$0.2800	128k
remote-deepseek-reasoner	$0.2800	$0.4200	131k
remote-deepseek-v4-flash-0731	$0.1400	$0.2800	1M
remote-gemini-2.0-flash-exp	—	—	—
remote-gemini-2.0-flash-thinking	—	—	—
remote-gemini-2.5-flash	$3.0000	$15.0000	1M
remote-gemini-2.5-flash-lite	$3.0000	$15.0000	1M

Modell	Input / 1M Token	Output / 1M Token	Kontext
remote-gemini-3.1-pro-preview	$3.0000	$15.0000	1M
remote-gemini-3.5-flash	$1.5000	$9.0000	1M
remote-gemini-3.5-flash-lite	$0.3000	$2.5000	1M
remote-gemini-3.6-flash	$1.5000	$7.5000	1M
remote-gemini-3.7-flash	$0.7500	$3.7500	1M
remote-gemini-3.8-flash	$0.7500	$3.7500	1M
remote-gemini-4-argon	$2.0000	$10.0000	1M
remote-gemma-4-31b-it	$0.1200	$0.3500	—
remote-gpt-4	$30.0000	$60.0000	8k
remote-gpt-4-turbo	$10.0000	$30.0000	128k
remote-gpt-4.1-mini	$0.4000	$1.6000	1M
remote-gpt-4.5-preview	—	—	—
remote-gpt-4o	$2.5000	$10.0000	128k
remote-gpt-4o-mini	$0.1500	$0.6000	128k
remote-gpt-5	$1.2500	$10.0000	272k
remote-gpt-5-mini	$0.2500	$2.0000	272k
remote-gpt-5-nano	$0.0500	$0.4000	400k
remote-gpt-5.2-codex	$1.7500	$14.0000	272k
remote-gpt-5.2-pro	$21.0000	$168.0000	272k
remote-gpt-5.3-codex	$1.7500	$14.0000	272k
remote-gpt-5.5	$5.0000	$30.0000	128k
remote-gpt-5.6-luna	$0.4000	$1.8000	1M
remote-gpt-5.6-sol	$4.0000	$20.0000	240k
remote-gpt-5.6-terra	$2.0000	$12.0000	240k
remote-gpt-6-astra	$10.0000	$50.0000	272k
remote-gpt-6-luna	$0.1000	$0.5000	64k
remote-gpt-6-sol	$2.0000	$10.0000	128k
remote-grok-4.5	$2.0000	$6.0000	500k

Modell	Input / 1M Token	Output / 1M Token	Kontext
remote-grok2-latest	—	—	—
remote-kimi-k2.5	$0.6000	$3.0000	262k
remote-kimi-k2.6	$0.9500	$4.0000	262k
remote-kimi-k3	$2.7590	$13.7900	1M
remote-kimi-k3.1-pro	$2.0000	$10.0000	1M
remote-ling-3.0-flash	$0.0210	$0.0630	262k
remote-minimax-m2.7	$0.3000	$1.2000	—
remote-minimax-m3	$0.3000	$1.2000	1M
remote-minimax-m3.1-flash	$0.1500	$0.6000	131k
remote-nex-n2-mini	$0.0250	$0.1000	—
remote-o1	$15.0000	$60.0000	200k
remote-o3	$1.0000	$4.0000	200k
remote-o4-mini	$0.5500	$2.2000	200k
remote-openrouter-deepseek-v3	$0.3000	$1.5000	—
remote-perplexity-sonar-deep-research	—	—	—
remote-perplexity-sonar-reasoning-pro	—	—	—
remote-qwen3.7-flash	$0.0300	$0.1300	—
remote-xai-grok-3	—	—	—
remote-xai-grok-3-mini	—	—	—
remote-xai-grok-4	—	—	—