Free LLM API Resources for Notice Board

Free Access to LLM APIs: A list of providers offering free or trial-based API access to Large Language Models. Please use these resources responsibly.

Provider	Models	Limits	Link
Groq	Various, including Distil Whisper Large V3 to Gemma 2 9B, Llama series	30 to 14,400 requests/day	https://console.groq.com
OpenRouter	Hermes 3 Llama, Liquid LFM 40B, Llama 3 series	20 requests/minute, 200 requests/day	https://openrouter.ai
Google AI Studio	Gemini 1.5 Flash, Gemini Pro, text-embedding models	1,500 requests/day	https://aistudio.google.com
Lambda Labs	Dracarys 2 72B, Liquid LFM 40B, Llama series	Requires credit card verification	https://docs.lambdalabs.com/on-demand-cloud/using-the-lambda-chat-c
Mistral	Mistral proprietary and open models	Free tier, phone number verification	https://console.mistral.ai / https://codestral.mistral.ai
HuggingFace	Various open models under 10GB	1,000 requests/day	https://huggingface.co/docs/api-inference/en/index
SambaNova Cloud	Llama 3.1, Llama 3.2 series	10 to 30 requests/minute	https://cloud.sambanova.ai
Cerebras	Llama 3.1 8B, Llama 3.1 70B	Waitlist, 30 requests/min	https://cloud.cerebras.ai
GitHub Models	Cohere, Meta-Llama, OpenAI GPT-4o, Phi 3 series	Variable limits based on subscription	https://github.com/marketplace/models
OVH AI Endpoints	CodeLlama, Llama 3 series, Mathstral	Token renews weekly, 12 requests/minute	https://endpoints.ai.cloud.ovh.net/
Cloudflare Workers AI	Llama series, Mistral, Gemma	10,000 tokens/day	https://developers.cloudflare.com/workers-ai
Together	Llama 3.2 11B Vision Instruct	Free for 2024	https://together.ai

Providers with Trial Credits

Together		
$5		
None	Various open models	https://together.ai
Fireworks		
$1		
None	Various open models	https://fireworks.ai
Unify		

$10 + $40 for contact			
None	Routes to other providers, various open and proprietary models	https://unify.ai	
DeepInfra			
$1.80			
None	Various open models	https://deepinfra.com	
NVIDIA NIM			
1,000 API calls for 1 month			
None	Various open models	https://build.nvidia.com/explore/discover	
AI21			
$10 for 3 months			
None	Jamba/Jurrasic-2	https://studio.ai21.com	
NLP Cloud			
$15			
Phone number verification	Various open models	https://nlpcloud.com/home	
Upstage			
$10 for 3 months			
None	Solar Pro/Mini	https://console.upstage.ai	
Baseten			
$30			
None	Any supported model - pay by compute time	https://app.baseten.co	
xAI			
$25/month			

until end of 2024		
None	Grok	https://console.x.ai
Hyperbolic		
$10		
None	Various models, including DeepSeek V2.5, Hermes Llama, palm-series	https://app.hyperbolic.xyz