

# OpenAI Proxy

Proxy Server to call 100+ LLMs in a unified interface & track spend, set budgets per virtual key/user

Features:

- **Unified Interface:** Calling 100+ LLMs Huggingface/Bedrock/TogetherAI/etc. in the OpenAI ChatCompletions & Completions format
- **Cost tracking:** Authentication, Spend Tracking & Budgets Virtual Keys
- **Load Balancing:** between Multiple Models + Deployments of the same model - LiteLLM proxy can handle 1.5k+ requests/second during load tests.

“ ???? LLM  
?????  
?? OpenAI Proxy ?????  
• ?? API ?????  
• ???  
• ???

- Doc: [https://docs.litellm.ai/docs/simple\\_proxy](https://docs.litellm.ai/docs/simple_proxy)