AI Agent Spending Limit Checklist: OpenAI, Claude and your AI agents

Set hard limits and control AI agent costs before they surprise you. OneSignalDesk · checked October 19, 2026 · free

An alert is not a cap. Tick a box only when the setting *blocks* requests.

1. Model provider

☐ **OpenAI, organization:** set a monthly spend limit and turn on the hard limit. Requests then return a 429 error once tracked spend reaches it.

☐ **OpenAI, project:** one project per agent, each with its own hard spend limit.

☐ **Anthropic, organization:** set your own monthly limit on the Billing page in the Claude Console.

☐ **Anthropic, workspace:** give each agent its own workspace (the Default Workspace can't take limits), then Workspaces → select it → **Spend limits**.

☐ **Claude paid plans, usage credits:** use “Adjust limit” to set a monthly spend limit instead of “Set to unlimited”, and decide on auto-reload on purpose.

2. Gateway (if you use one)

☐ **Cloudflare AI Gateway:** select your gateway → Settings > Spend limits → add a rule. Split by a metadata key to give each agent its own budget (up to 20 rules per gateway).

☐ **OpenRouter:** set a credit limit on every API key.

☐ **Self-hosted LiteLLM Proxy:** run 1.83.9 or later (CVE-2026-59823).

3. Automation tool

☐ **Zapier:** decide on pay-per-task. With it off, Zaps stop at your plan's task limit. With it on, overage is billed and Zaps stop at 3 times your plan's limit.

☐ **Zapier AI steps:** keep the per-run task limit (75 by default, up to 500). A live run that reaches it pauses for your approval.

☐ **Make:** leave **Extra credits auto-purchasing** off, so “your scenarios won't continue to run” when credits run out.

☐ **n8n:** set Max Iterations on the AI Agent node. Gateway credits (n8n Cloud) stop at a zero balance; with your own keys, cap the key at the provider (section 1 or 2).

4. Hygiene

☐ Set alerts below each cap (for example 50% and 80%), so the cap is never your first warning.

☐ Set caps **below** what you can afford to lose. OpenAI says enforcement isn't instantaneous.

Sources and method. Vendors' own pages, checked October 19, 2026: OpenAI spend limits (developers.openai.com/api/docs/guides/spend-limits) · Anthropic Workspaces and Rate limits (platform.claude.com/docs) · Claude usage credits (support.claude.com/en/articles/12429409) · Cloudflare AI Gateway spend limits (developers.cloudflare.com/ai-gateway/features/spend-limits) · OpenRouter authentication (openrouter.ai/docs) · Zapier pay-per-task and AI by Zapier help pages · Make extra credits (help.make.com/extra-credits) · Make pricing FAQ (make.com/pricing) · n8n AI Agent node, Tools Agent (Max Iterations; docs.n8n.io) · n8n Gateway credits (docs.n8n.io) · CVE-2026-59823 (cve.org). We haven't tested these settings ourselves. No affiliate links. Full guide: onesignaldesk.com/guides/ai-agent-spending-caps