Sub2API ??????????????? 2026.06?

Sub2API 风控中心功能调研: omni-moderation-latest 模型速率限制、延迟影响、审核效果评估

- Sub2API 内容审计（风控中心）功能说明（2026.06）

Sub2API ??????????????? 2026.06?

> 数据来源: LINUX DO 社区帖子逐楼层阅读 + OpenAI 官方文档, 2026-06-12 整理。覆盖 7 个相关帖子、50+ 楼层。

??????

Sub2API 的「风控中心」功能在请求转发到上游前，使用 AI 模型检测高风险内容（逆向、破限、Cyber Abuse 等），拦截危险请求以保护上游账号不被封禁。

- 启用方式: 系统设置 → 功能开关 → 风控中心 → 启用
- 默认模型: OpenAI `omni-moderation-latest`（官方专用内容审核模型）
- 模型费用: 免费，但需要 OpenAI 官方 API Key + 充值至少 $5 解锁
- 附加功能: 2026-05-20 起新增「关键词检测」，可自定义敏感词列表

> 来源: 2129973 @samhuang 1楼（2026-05-07）; 2175491 @yuancjun 2楼（2026-05-14）

??????

`omni-moderation-latest` 模型本身的限制是 账号级别（同一账号下多个 Key 共享同一限制池）：

限制项	数值	说明
RPM (每分钟请求数)	500	账号级，多 Key 共享
TPM (每分钟 Token 数)	10,000	每次审核请求消耗约 100~500 token
RPD (每天请求数)	10,000	自用足够，多人共享可能紧张
TPD (每天 Token 数)	1,000,000	充裕

?????????

- **Free Tier** (未充 **$5**) : 并发仅 3~5 个, 几乎立即触发 429
- 充 **$5** 后: RPM 提升但仍有限制, 一个账号开多个 Key 无法突破账号级限制
- 多 **Key** 无效: 限制是账号级别的, 同一账号多 Key 共享 RPM/TPM 池
- 解决方案: 如需更高并发, 需要多个 OpenAI 账号各开独立 Key

> 来源: 2177401 @dark1222 2楼 (2026-05-14) 确认官方数值; 2125872 @yuaotian86 6楼 (2026-05-26) : 「免费级别一个号并发就 3~5 个」; 2215661 @Zooe 1楼 (2026-05-20) : 「充了 5 刀用了官方 key 但老是冻结, 一个账号开多个 key 没用」

??????

开启风控中心后, 每个请求都需要先经过审核模型判断, 通过后才转发到上游, 这会增加额外延迟。

环节	预估耗时	说明
omni-moderation 调用	100~500ms	轻量分类模型, 取决于输入长度
网络往返 (sub2api ↔ OpenAI)	200~800ms	取决于服务器到 OpenAI API 的网络延迟
审核结果判断 + 转发	50~100ms	Sub2API 本地处理
总延迟增加	**0.5~2** 秒	首字延迟 (TTFT) 增加明显

影响最大的场景: 流式请求 (Streaming) 需要等审核通过后才开始返回数据, 用户会感受到明显的「思考时间」变长。

> 来源: 2175491 @Mxucc 13楼 (2026-05-15) : 「加这个延迟会非常高吧, 毕竟要先审计一下」; @Orange4664 14楼: 「对的确实高了点, 没办法, 孩子害怕」

????????

??????

- 仇恨言论（ hate）
- 暴力内容（ violence）
- 自残相关内容（ self-harm）
- 色情内容（ sexual）
- 骚扰 /威胁（ harassment）

???????

- 逆向工程请求 （如「请逆向一下这个 app」） —— 检测不出
- 破解限制（破限） —— 检测不出
- 网络安全攻击类 **prompt** —— 检测不出
- 伪装后的恶意请求 —— 容易绕过

结论： omni-moderation-latest 对常规有害内容有效，但对 Codex/ChatGPT 场景下的高风险行为（逆向、破限、 Cyber Abuse）基本无效 。这些行为需要依赖 OpenAI 自身的 5.5 风控机制来识别。

> 来源： 2215661 @y1285012 4楼（ 2026-05-20）：「官方 omni-moderation-latest 这个检测没啥用，我试过发送：请逆向一下这个 app，都检测不出来，但是这个 prompt 使用 gpt5.5，会触发 cyber 验证的」； 2129973 @apparition 12楼：「我其实很怀疑 2024-09-26 推出的模型能侦测到什么」

????????????

2026-05-20 起， Sub2API 新增「关键词检测」功能，可自定义敏感词列表作为 omni-moderation 的补充。

- 优势： 不依赖外部 API，无额外延迟，无速率限制
- 劣势： 需要自行维护关键词列表，无法识别语义变体
- 适用场景： 拦截明确的逆向 /破限关键词，比 omni-moderation 更实用

> 来源： 2215661 @mollys 7楼（ 2026-05-20 ）：「现在有关键词检测的功能了，可以开下试一试」

??????

配置项	要求	说明
API Key 类型	OpenAI 官方 Key	中转站 /公益站不支持 omni-moderation
最低充值	$5	不充值只有 Free Tier，几乎不可用
Key 来源	无订阅的纯 API 账号	有订阅的号开 API 可能被封
充值有效期	每年需充 $5	余额过期后 API 权限降级

> 来源: 2175491 @apparition 5楼: 「不要用有订阅的号开 api，可能会封」；@yb3712590 16楼: 「 openai 官网的充值是有有效期的，每年都需要充 5 美元」；2143094 @timlong 9楼: 「新注册的 free 账户创建的 key，一放到 sub2api 风控中心里面，测试就报 429，然后冻结」

??????

维度	结论
费用	模型免费，需充 $5 解锁 API
RPM	500（账号级，多 Key 共享）
RPD	10,000（自用够，多人共享紧张）
延迟	增加约 0.5~2 秒首字延迟
检测效果	对 hate/violence/sexual 有效，对逆向 /破限基本无效
推荐做法	omni-moderation + 关键词检测 组合使用

??L ?????

帖子 ID	主题	日期
2129973	sub2api 风控中心如何使用？（功能介绍核心帖）	2026-05-07
2125872	有没有佬有 omni-moderation-latest 这个审计模型的渠道	2026-05-07
2177401	请问有哪家支持这个 omni-moderation-latest 模型的（含速率限制数据）	2026-05-14
2175491	sub2api 可以看到哪个用户触发了 cyber abuse 嘛（含延迟讨论）	2026-05-14
2143094	求一个 OpenAI 官 key，只用 omni 模型做风控管理	2026-05-09
2215661	关于 sub2 内容审计设置问题（含审核效果实测 + 关键词检测）	2026-05-20

帖子 ID	主题	日期
1993856	cpa 和 sub2api, 测速对比 (含 sub2api 延迟基准数据)	2026-04-18