法律与合规 03

协议模板/草案

仅供审阅，以最终签署版本为准

AI 服务政策

模型准入、内容安全、敏感词控制、用量记录与违规处理规则。

生效日期	2026 年 6 月 29 日
最后更新	2026 年 7 月 23 日
版本	2026-07-23

政策目录

能力边界核对（2026-07-23）：应用支持并默认启用 Prompt 敏感词检查，具体环境是否启用及词库范围取决于管理员配置。本政策只描述现行行为规则和已验证的应用能力；后续新增实质性控制时将更新政策并另行通知。

1. 政策概述

Dogzee Technology Limited（二狗子科技有限公司） 运营二狗子 AI API 服务平台。本政策说明我们如何：

- 管理 AI 模型和服务的可用性
- 执行内容安全控制和敏感词屏蔽
- 记录用量数据并保护用户隐私
- 处理违规行为和投诉

本政策是用户协议的重要补充，使用我们的服务即表示您同意遵守本政策。

2. 服务范围

2.1 我们提供的服务

- **AI 模型 API 接入**：支持主流 AI 模型（如 OpenAI、Anthropic、Google 等）
- **兼容接口**：提供与 OpenAI API 兼容的端点
- **请求路由**：智能分发请求至最优渠道
- **密钥管理**：创建和管理 API Key
- **用量统计**：实时 Token 计量和消耗记录
- **计费系统**：额度充值、订阅套餐、账单管理

2.2 服务提供方式

我们的 AI 服务可能通过以下方式提供：

- 自有基础设施
- 合作伙伴渠道
- 第三方模型服务商

具体提供方式取决于：模型可用性、价格、合规要求、用户配置和平台策略。

2.3 服务调整权利

基于以下原因，我们可能**随时**新增、移除、暂停或限制模型、渠道或功能：

- 安全风险或滥用行为
- 法律法规或合规要求
- 第三方服务变更或中断
- 商业或运营考虑
- 风控和反欺诈需要

我们会尽量提前通知，但紧急情况下可能无法提前告知。

3. 模型与产品上架标准

为保障平台合规运营，我们对 AI 模型和渠道执行以下上架审查。

3.1 上架前评估

在向用户开放任何模型、渠道或 AI 能力前，我们会进行**多维度审查**：

评估维度和审查内容：

- **模型来源**：模型提供商是否合法合规、信誉良好
- **功能类型**：支持的输入输出类型（文本、图像、代码等）
- **内容安全**：模型是否有内置安全控制、是否容易被滥用
- **合规性**：是否符合香港及主要市场的法律要求
- **隐私保护**：数据处理方式、是否涉及敏感信息训练
- **稳定性**：服务可用性、响应速度、错误率
- **定价**：成本是否合理、是否支持我们的计费模式
- **地区限制**：是否有地域访问限制或出口管制

3.2 上架决策

如果模型或能力存在以下风险，我们会**拒绝上架或立即下架**：

- 存在重大法律风险或违反监管要求
- 缺乏有效的内容安全控制
- 容易被用于违法或有害用途
- 涉及未授权的数据训练或侵权

- 隐私保护措施不足
- 服务稳定性或安全性无法保障
- 计费或结算存在欺诈风险

3.3 持续监控

已上架的模型会持续接受监控。如发现以下情况，我们会采取限制或下架措施：

- 滥用率异常升高
- 收到大量投诉或违规报告
- 第三方服务条款变更
- 新的法律法规要求

4. 用户责任与禁止用途

4.1 用户的责任

您对以下内容**承担全部责任**：

- 您提交的提示词（Prompt）和上传内容
- API 集成的应用场景
- 模型输出内容的下游使用
- 对最终用户的合规告知

4.2 严格禁止的用途

严禁将我们的服务用于以下场景：

违法犯罪

- 生成违法内容（如诈骗话术、假证件）
- 协助洗钱、恐怖融资、走私等犯罪活动
- 侵犯知识产权（如批量生成盗版内容）

有害内容

- 暴力、血腥、虐待内容
- 色情或涉及未成年人的不当内容
- 仇恨言论、歧视或煽动性内容
- 自残、自杀相关的引导性内容

侵犯权益

- 未经授权冒充他人身份
- 侵犯隐私（如人肉搜索、信息泄露）
- 骚扰、威胁、跟踪他人

恶意技术行为

- 制作或传播恶意软件、病毒
- 攻击网络或系统安全
- 绕过平台安全控制或技术限制

滥用与欺诈

- 批量生成垃圾信息或虚假评论
- 操纵舆论或虚假宣传
- 学术作弊（如代写论文）
- 其他违反公序良俗的行为

4.3 合规要求

您必须确保使用行为符合：

- 适用的法律法规（包括但不限于香港法律）
- 行业规范和职业道德
- 特定模型提供商的附加条款（如 OpenAI Usage Policy）

5. 内容安全与敏感词控制

内容安全与敏感词控制是平台合规运营的重要机制。

5.1 应用支持的检测机制

应用支持并默认启用敏感词检查：使用已配置的敏感词表和 Aho-Corasick 精确匹配扫描已提取的用户提示词。匹配前会统一文本大小写，但这不等同于模糊匹配、变体识别或上下文意图分析。

具体环境是否启用、词库覆盖范围、更新频率和检测效果取决于管理员配置。

5.2 检测范围

应用支持的范围：

- **用户提示词**：经过已启用的敏感词检查。
- 实际覆盖范围取决于请求格式的文本提取逻辑；不能据此推断所有参数、附件或嵌套字段

都已被扫描。

5.3 请求拦截流程

管理员启用敏感词检查后，应用执行 Prompt 检测和请求侧拦截：

```
用户请求 → 敏感词检测 → 发现违规 → 立即拦截
                ↓
返回错误提示（不执行实际 API 调用）
```

拦截响应示例：

```
{
 "error": {
 "message": "请求包含敏感内容，已被拦截",
 "type": "new_api_error",
 "code": "sensitive_words_detected"
 }
}
```

5.4 违规处理措施

平台可依据用户协议、人工复核、实际风险和适用法律限制 API Key 或账号。措施可能包括请求拦截、限流、暂停或终止服务；具体处置会结合事件严重程度、证据和申诉结果判断，不承诺由系统按固定次数自动执行递进处罚。

5.5 人工复核

对于以下情况，我们可能进行人工复核：

- 自动检测不确定的边缘案例
- 用户申诉自动拦截结果
- 收到外部投诉或举报
- 监管部门要求调查

人工复核承诺：

- 仅查看违规内容片段和上下文
- 不会泄露或另作他用
- 复核结果会通过邮件通知

未来如增加输出侧审核、上下文分析或自动处罚等实质性控制，平台会先完成实现与验证，并按本政策的变更条款另行通知。

6. 提示词与响应内容处理

6.1 处理原则

我们承诺：默认不存储提示词和响应内容。

这些内容仅在以下必要场景**临时**处理：

- 完成您的 API 请求（转发给第三方模型）
- 执行内容安全检测（敏感词扫描）
- 排查技术故障（经用户授权）
- 应对安全威胁或滥用行为
- 响应合法的司法或监管要求

6.2 例外情况

以下情况可能由授权人员按事件需要最小化保留内容片段：

- **安全或滥用事件**：经人工判断需要审计或处理申诉时，记录必要的脱敏摘要

- **用户主动报告问题**：经用户同意保留问题示例，用于排查和改进
- **法律要求**：根据司法或监管要求保留必要证据

当前系统没有实现“每次命中自动保存违规内容并在 90 天后自动删除”的完整控制。具体保留范围和期限应依据事件、适用法律和已验证的删除能力决定。

6.3 第三方处理

您的请求会转发至第三方 AI 模型提供商（如 OpenAI、Anthropic 等）。请注意：

- 这些提供商有各自的数据处理政策
- 我们无法控制第三方如何使用您的数据
- 建议避免在提示词中包含敏感个人信息

7. 用量日志与 Token 消耗记录

7.1 记录的元数据

为了计费、排障、安全和风控，我们会记录：

记录的数据项及用途：

- **请求时间**：计费、排障
- **模型名称**：计费、统计
- **API Key 标识**：计费、溯源
- **用户账号**：计费、风控
- **输入 Token 数量**：计费依据
- **输出 Token 数量**：计费依据
- **额度消耗**：账单生成
- **请求状态（成功/失败）**：排障、质量监控
- **请求延迟**：性能优化
- **IP 地址（可选）**：风控、防刷

7.2 用户可查询的记录

您可以在控制台实时查看：

- 每次请求的详细日志（时间、模型、Token、费用）
- 按日期、模型、API Key 筛选的用量统计
- 额度扣减明细
- 订单和账单记录

这些记录可作为 Token 消耗的官方证明，用于对账或审计。

7.3 数据保留期限

- 用量日志和计费记录：保留 **90 天**
- 订单和账单记录：保留 **3 年**（会计和税务要求）

8. 投诉与违规处理

8.1 投诉渠道

如您发现其他用户滥用我们的服务，请通过以下方式举报：

- 支持邮箱：support@ergouzi.life
- 说明：违规账号或 API Key、违规内容描述、证据截图

8.2 调查流程

收到投诉、风险通知、支付机构要求或法律请求后，我们会：

1. **启动调查**：审查相关账号的请求记录和违规日志
2. **评估风险**：判断违规严重程度和影响范围
3. **采取措施**：根据调查结果执行处罚
4. **通知结果**：向举报人和当事人反馈处理结果（必要时）

8.3 处罚措施

根据违规性质，我们可能采取：

- 警告通知（邮件或控制台消息）
- 限制 API 请求速率或频次
- 暂时或永久禁用 API Key
- 暂停或关闭账号
- 拒绝后续注册或充值请求
- 保留违规证据配合法律调查
- 向监管部门或执法机关报告

8.4 申诉机制

如您认为处罚有误，可以申诉：

- 点击售后联系我们
- 或者发送邮件至：support@ergouzi.life
- 提供：账号信息、处罚通知、申诉理由、相关证据
- 我们会在 **3-5 个工作日**内复核并回复

申诉期间，处罚不会暂停执行，除非复核结果支持您的申诉。

9. 免责声明

9.1 第三方模型

我们提供的是 **API 聚合服务**，不是 AI 模型的开发者或训练者。对于第三方模型的输出内容、准确性、可靠性，我们不承担责任。

9.2 用户行为

您对通过我们的服务生成的内容承担全部责任。我们仅提供技术工具，不对您的使用场景或下游应用负责。

9.3 服务中断

尽管我们会尽力保障服务稳定，但以下情况导致的服务中断或数据丢失，我们不承担责任：

- 第三方模型提供商故障
- 不可抗力（自然灾害、战争、政府行为等）
- 网络攻击或安全事件
- 用户违规导致的账号封禁

10. 政策更新

我们可能不定期更新本政策以反映：

- 新的法律法规要求
- 服务功能调整
- 风控策略优化

重大变更会通过邮件或控制台通知。更新后继续使用服务即视为接受新版政策。

11. 联系我们

如对本政策有任何疑问，或需要举报违规行为，请联系：

- **支持邮箱**：support@ergouzi.life
- **响应时间**：工作日 1-3 个工作日

生效日期：2026 年 6 月 29 日

最后更新：2026 年 7 月 23 日

版本：2026-07-23