

可信大模型分控策略网关

用户名

admin

密码

登录



系统概览

网络配置

定向加速

智能组网

API网关

网络诊断

设备信息

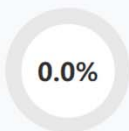
重启系统

系统概览

自动刷新：开启

运行时间：6天 19时 45分

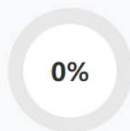
CPU 温度：34.0 °C



CPU 使用率



内存使用率
已用413M / 总1913M



存储使用率
已用6M / 总11382M

网口链路状态：



WAN
1000M



LAN1



LAN2
1000M



LAN3

网络信息

重启系统

网口链路状态:



LAN1



LAN2
1000M



LAN3



WAN
1000M

网络信息

LAN IP: 192.168.137.1

WAN 协议: dhcp

2.4G WiFi: 信道: 6

5G WiFi: 信道: 36

DHCP 已分配设备

192.168.137.125 00:e0:4c:68:3d:d6 Thinkpad-wzX1	
---	--

无线接入设备

系统概览

网络配置

定向加速

智能组网

API网关

网络诊断

设备信息

重启系统

账号与安全

账号安全与防护 (当前账号: admin)

修改当前登录账号 (admin) 密码

新密码

如需修改密码请在此输入 (至少6位)



确认新密码

再次输入新密码



防爆破安全控制

允许连续输错次数

0

设置为 0 表示不生效 (禁用锁定策略) ; 设置大于 0 时达到次数将锁定账号

触发锁定后限制时长 (分钟)

1

锁定时间内禁止登录

HTTPS 加密传输

☐ 启用 HTTPS 加密访问

启用后系统将通过加密通道 (HTTPS) 提供 Web 管理服务; 取消勾选将恢复标准 HTTP 访问

HTTPS 服务端口

443

系统概览

网络配置

定向加速

智能组网

API网关

网络诊断

设备信息

重启系统

192.168.10.130/rulenet

LAN 局域网配置

IP 地址

192.168.137.1

子网掩码

255.255.255.0

DHCP 服务

开启

WAN 上网配置

IP/mask192.168.10.130/24

网关192.168.10.1

DNS192.168.10.1 114.114.114.114

上网协议

DHCP (自动获取)

路由器工作模式

工作模式

旁路模式

路由模式: 正常上网, WAN口作为外网入口

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

故障切换与负载策略

负载模式: 主备模式 (按服务商列表优先级顺序首选, 故障顺延切换) 故障隔离时长: 连接异常临时禁用 1800 秒 保存策略配置

说明: 1. 故障隔离时长同时作用于「内网分发多服务商直连」与「模型池调度」, 故障节点在隔离期满后由后台自动探测恢复; 2. 负载模式控制内网客户端勾选了多个服务商时的选路方式 (若处于主备模式, 可通过服务商列表的 ▲ / ▼ 按钮调整主备优先级顺序; 模型池内部始终按池编排权重负载均衡)。

第三方 AI 服务提供商 (上游)

一键清零统计

服务商名称 接口协议规范 第三方 API Base URL 第三方真实 API Key (可选) 添加上游

优先级排序	服务商 (协议)	API URL	真实 API Key	用量与词元(Tokens)	操作
1	天鸿智算 OPENAI 异常禁用中	https://api.thhtcloud.com	sk-*****KVwu 查看	总请求: 3 次 总 Tokens: 0	限定模型 重置 删除
2	博伦智汇 OPENAI	https://max.booleanint.co...	sk-*****4LVI 查看	总请求: 4 次 总 Tokens: 0	限定模型 重置 删除
3	硅基流动 OPENAI	https://api.siliconflow.cn	sk-*****mxif 查看	总请求: 0 次 总 Tokens: 0	限定模型 重置 删除
4	阿里云百炼 OPENAI	https://ws-cooirs8qgk0jsp...	sk-*****DYZQ 查看	总请求: 19 次 总 Tokens: 143,019	限定模型 重置 删除

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

负载模式:

主备模式 (按服务商列表优先级顺序首选, 故障顺延切换) ▼

故障隔离时长: 连接异常临时禁用

1800

秒

保存策略配置

说明: 1. 故障隔离时长同时作用于「内网分发多服务商直连」与「模型池调度」, 故障节点在隔离期满后由后台自动探测恢复;
2. 负载模式控制内网客户端勾选了多个服务商时的选路方式 (若处于主备模式, 可通过服务商列表的 ▲ / ▼ 按钮调整主备优先级顺序; 模型池内部始终按池编排权重负载均衡)。

第三方 AI 服务提供商 (上)

服务商名称

例如: Claude 官方

接口

优先级排序

服务商 (协议)

1 ▲ ▼

天鸿智算
OPENAI

2 ▲ ▼

博伦智汇
OPENAI

3 ▲ ▼

硅基流动
OPENAI

4 ▲ ▼

阿里云百炼
OPENAI

5 ▲ ▼

千帆
OPENAI

限定模型设定 - 阿里云百炼

保存限定模型

×

点击模型切换限定状态 (蓝色为启用, 灰色为停用, 全灰为不限定模型):

cmaas-qwen3.8-flash

qwen-image-2.1-pro

qwen3.8-omni-flash-realtime

stepfun/step-5-preview

glm-5.3-prime

ZHIPU/GLM-5.3-FlashX

vanchin/deepseek-v4.1-flash

qwen-audio-3.1-realtime-plus

qwen-mt-uni

qwen3.8-omni-flash

qwen3.8-livetranslate-flash-realtime

kimi/kimi-k2.8-preview

glm-5.3

deepseek-v4.1-flash

qwen3.8-max-0902

qwen3.7-text-embedding-flash

一键清零统计

Key (可选)

添加上游

用量与词元(Tokens)

操作

/wu 查

总请求: 3 次
总 Tokens: 0

限定模型

重置

删除

VI 查看

总请求: 4 次
总 Tokens: 0

限定模型

重置

删除

if 查看

总请求: 0 次
总 Tokens: 0

限定模型

重置

删除

sk-*****DYZQ 查
看

总请求: 19 次
总 Tokens: 143,019

限定模型

重置

删除

https://ws-cooirs8qgk0jsp...

看

bce*****c342 查
看

总请求: 3 次
总 Tokens: 0

限定模型

重置

删除

https://qianfan.baidubce....

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

OPENAI		看		总 Tokens: 143,019				
5	▲▼	千帆	https://qianfan.baidubce....	bce*****c342 查	总请求: 3 次 总 Tokens: 0	限定模型	重置	删除
OPENAI		看						
6	▲▼	UCloud	https://api.modelverse.cn	90i*****F385 查	总请求: 76 次 总 Tokens: 1,103,714	限定模型	重置	删除
OPENAI		看						
7	▲▼	UCloud2	https://api.modelverse.cn	cE*****1d0e 查	总请求: 72 次 总 Tokens: 553,280	限定模型	重置	删除
OPENAI		看						

模型池管理

模型池名称

模型名

例如: 旗舰GPT池

例如: gpt-4o

新建模型池

模型池名称	匹配模型名	上游路由状态	操作
111	glm-5.2	已挂载 2 个服务商	调度编排 删除池
2222	glm	已挂载 0 个服务商	调度编排 删除池
test	deepseek-v4.1-flash	已挂载 4 个服务商	调度编排 删除池

- 系统概览
- 网络配置
- 定向加速
- 智能组网
- API网关
- AI服务商
- 内网分发
- API网关高级
- 图表统计
- 网络诊断
- 设备信息
- 重启系统

模型池调度编排 : 111 (目标匹配: glm-5.2)

智能扫描匹配服务商

加入池	上游服务商	并发负载策略(权重1-10)	每日Token限额(0为不限)	今日已用 Token
☒	天鸿智算 OPENAI 异常禁用中	1	0	0
☐	博伦智汇 OPENAI	1	0	0
☐	硅基流动 OPENAI	1	0	0
☐	阿里云百炼 OPENAI	1	0	0
☒	千帆 OPENAI	1	0	0
☐	UCloud OPENAI	1	0	0
☐	UCloud2 OPENAI	1	0	0

保存模型池编排配置 返回列表

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

分发内网用户 API Key (支持单 Key 绑定多服务商/多模型池自动路由)

一键清零展示累计数据

接口协议规范:

OpenAI 兼容协议

绑定目标 (可勾选多个):

★ 111

所属工作组:

默认

自定义/生成的 Key:

070a765df0b31809344d48a82f63b2f9

组成员:

例如: 张三

添加分发 Key

提示: 当一个内网客户端勾选了多个服务商或模型池时, 客户端拉取模型列表将自动合并所有目标模型, 并在提问时根据所选模型名称自动匹配对应的服务商转发。

工作组: 默认 (5 个账号)							组累计请求: 122	组累计 Tokens: 1,257,566
内网分发 Key (客户端)	接口协议	映射目标 (可多选池/商)	所属工作组	组成员	累计展示数据	操作		
b262bb*****3a55 查看	OpenAI 兼容协议	博伦智汇	默认	小茜2	累计请求: 2 次 累计 Token: 0	重置 删除		
dfe31d*****9b1a 查看	OpenAI 兼容协议	★ 111	默认	小茜	累计请求: 3 次 累计 Token: 0	重置 删除		
59865a*****4107 查看	OpenAI 兼容协议	UCloud	默认	老王	累计请求: 71 次 累计 Token: 963,744	重置 删除		
1dee9e*****80e7 查看	OpenAI 兼容协议	UCloud2	默认	666	累计请求: 46 次 累计 Token: 293,822	重置 删除		
bbad70*****3132 查看	OpenAI 兼容协议	★ 111	默认	6661	累计请求: 0 次 累计 Token: 0	重置 删除		

工作组: 张三组 (1 个账号)							组累计请求: 87	组累计 Tokens: 551,846
------------------	--	--	--	--	--	--	-----------	---------------------

- 系统概览
- 网络配置
- 定向加速
- 智能组网
- API网关
- API网关高级
- API网关统计
- 网络诊断
- 设备信息
- 重启系统

工作组管理

请输入新工作组名称

添加工作组

工作组名称 (修改后联动同步)	组成员总数	操作
默认	9 人	默认组不可删
项目开发	3 人	删除组
项目开发二组	0 人	删除组

保存工作组配置

工作组 Token 配额策略

工作组	全组每日总配额	全组每月总配额	组成员单日配额	组成员单月配额
默认	已用: 15 上限: 10000	已用: 19,735 上限: 100000	最高已用: 15 上限: 0	最高已用: 19,720 上限: 0
项目开发	已用: 0 上限: 1000	已用: 0 上限: 0	最高已用: 0 上限: 0	最高已用: 0 上限: 0
项目开发二组	已用: 0 上限: 0	已用: 0 上限: 0	最高已用: 0 上限: 0	最高已用: 0 上限: 0

组成员 Token 配额策略

添加/设置组成员专属策略:

选择或输入组成员: 输入关键字模糊匹配

每日配额: 0

每月配额: 0

添加策略

组成员	所属工作组	每日专属 Token 配额	每月专属 Token 配额	操作
小洋洋	默认	已用: 0 上限: 1000	已用: 0 上限: 10000	移除策略
杨洋	项目开发	已用: 0 上限: 2000	已用: 0 上限: 10000	移除策略

配额填 0 表示继承工作组默认策略。修改后请点击下方保存按钮。

保存组成员配额策略

Token 超额告警与阻断策略

超额告警提示

开启客户端超额告警

超额阻断机制

只提示不阻断 (先回复告警内容, 再连接AI答案)

▼

1.

▼

7

可信大模型分控策略网关

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

API 网关服务状态 (支持 OpenAI / Claude / Gemini 统一协议自动转换)

当前网关运行状态: **运行中 (HTTP)** (授权上限: **6 / 无限制** 用户)

内网发布服务地址与端口:

http://0.0.0.0:3000

保存配置

停止服务

💡 提示: 在「内网分发」页面可为每个 Key 勾选多个服务商或模型池, 网关会根据客户端请求的模型自动匹配对应服务商并完成双向协议转换。

双向安全围栏与 DLP 动态脱敏防护

检测方向管控

☒ 启用输入检测 (提问阻断 / 发送前动态脱敏) ☒ 启用输出检测 (回答熔断 / 流式回答脱敏)

 **内置隐私数据识别规则 (DLP)**

启用此规则



☒ 二代身份证号 ☒ 手机号码 ☒ 银行卡号 ☒ 企业统一信用代码 ☒ 电子邮箱

☒ API 密钥凭证 (sk-/AKIA/AIza/ghp_) ☒ JWT 认证令牌 ☒ 数据库连接串

☒ 私钥证书块 (PEM/RSA) ☒ 内网 IPv4 地址 ☒ 互联网 IPv4 地址 ☒ MAC 物理地址

处置模式:

动态脱敏模式 (自动将隐私信息替换为 *** 掩码后放行)

可信大模型分控策略网关

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

动态脱敏模式 - 敏感词与正则列表

启用此表

命中此表自动替换为 *** 掩码并继续正常请求大模型（一行一条，支持 regex:表达式）

内部代号A
客户真实姓名
regex:password\s*[:=]\s*\S+

严格阻断模式触发后的自定义拦截提示内容

⚠ 您的提问或 AI 的回答包含不合规内容，已被系统拦截。

保存安全围栏配置

轻量级智能问答缓存（降低 Token 消耗）

一键清空缓存与统计

当前持久化缓存条目：8 条 | 累计为企业节省请求（命中次数）：1 次

智能问答缓存开关

启用智能问答缓存

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

组成员接入安全（IP 与设备 MAC 地址准入控制）

添加组成员安全设备绑定（支持单个成员绑定多个 IP 和多个 MAC）：

选择或输入组成员： 如：张三

绑定访问 IP： 如 192.168.1.10, 192.168.

绑定设备 MAC： 如 aa:bb:cc:11:22:33

+ 添加到绑定列表

接入安全验证开关：

启用设备 IP/MAC 准入校验

未授权设备拦截提示：

⚠ 您使用了未授权的设备进行访问，请联系管理员。

组成员	所属工作组	允许访问的 IP 列表 (逗号分隔)	允许访问的设备 MAC 列表 (逗号分隔)	操作
老王	默认	192.168.30.186	多个MAC用逗号分隔，留空不限MAC	移除绑定

💡 说明：一旦为某成员配置了绑定记录，当其数据包的访问 IP 或通过 ARP 解析出的设备 MAC 地址不在允许列表内时，网关将立即阻断访问，并记录类型为「未授权访问」的报警日志。

保存组成员接入安全配置

远程安全日志审计

远程日志审计开关

启用 Webhook 异步审计日志推送

审计粒度模式

完全记录日常所有对话提问与回答全文

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

远程安全日志审计

远程日志审计开关

启用 Webhook 异步审计日志推送



审计粒度模式

完全记录日常所有对话提问与回答全文



远端 HTTP Webhook 接收地址

http://172.16.1.49:3000/webhuuk/test

Webhook 鉴权 Token (可选)

磁盘缓存

关闭磁盘缓存 (推送异常直接丢弃新日志)



磁盘缓存路径 (绝对路径)

磁盘缓存最大空间限制 (MB)

50

系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

图表统计

报警日志

⚙️ 数据保存间隔配置: 24 小时 (默认) (控制数据的聚合采样及报表输出颗粒度)

📁 数据最长保存时间: 1 年 (365天 / 默认) [保存配置](#)

数据概览

今天

昨天

本月

近7天

近30天

总请求次数

11 次

Token 总消耗

52,245

安全与准入拦截数

0 次

Token 超额告警

0 次

自定义多维图表与报表输出

📄 下载报表 (CSV)

图表样式:

柱状图 (Bar)

开始时间:

2026-10-15 00:00

结束时间:

2026-10-15 23:59

统计指标:

- ☐ 请求次数
☐ 消耗 Token
☐ 拦截次数
☐ 超额告警

服务商:

搜索服务商...

- ☐ 天鸿智算
☐ 博伦智汇
☐ 硅基流动

工作组:

搜索工作组...

- ☐ 张三组
☐ 默认

网络诊断

系统概览

网络配置

定向加速

智能组网

API网关

网络诊断

设备信息

重启系统

Ping 连通性测试

目标地址

www.baidu.com

测试次数

4

开始测试

点击“开始测试”执行Ping连通性检测

Traceroute 路由追踪

目标地址

www.baidu.com

开始追踪

点击“开始追踪”执行路由追踪

Nslookup 域名解析

域名

www.baidu.com

开始解析

点击“开始解析”执行域名解析

10:46

2026/9/23



系统概览

网络配置

定向加速

智能组网

API网关

- AI服务商

- 内网分发

- API网关高级

- 图表统计

网络诊断

设备信息

重启系统

系统在线升级

当前系统版本: **v1.1.0**

在线检查新版本

💡 提示: 升级过程中设备可能会短暂断开连接, 完成后将自动重启相关服务。

设备基本信息

设备型号

EDGE Toker50R

设备序列号 (SN)

84e2a618582d55f128a7d2885107eaf0

公司名称

常州万联云网信息技术有限公司

联系电话

400-0277-590

定向加速

网络配置

定向加速

智能组网

API网关

网络诊断

设备信息

重启系统

定向加速管理

服务开关

当前运行: 运行中

开启=立即运行并开机自启 | 关闭=立即停止并禁止自启

选择节点

美国

选择后会同时应用到TCP和UDP协议

加速模式

分流模式(默认)

分流模式只走指定规则流量，全局模式所有流量走代理节点

DNS防泄漏

开启后访问国内网络可能会有影响

直连列表

代理列表

屏蔽列表

保存并应用

当前定向出口流量

上传流量: 0.34 GB

下载流量: 0.64 GB

总计流量: 0.98 GB

14:30

2026/9/23

重启系统

加速模式

分流模式(默认)

分流模式(默认)

全局模式

开启后访问国内网络可能会有影响

直连列表

代理列表

屏蔽列表

保存并应用

当前定向出口流量

上传流量: 0.34 GB

下载流量: 0.64 GB

总计流量: 0.98 GB

当前定向出口IP

IP: 38.34.18.109

城市: Los Angeles

省份: California

国家: 美国

运营商: AS174 Cogent Communications, LLC

查看更多

可信大模型分控策略网关

- 系统概览
- 网络配置
- 定向加速
- 智能组网
- API网关
- 网络诊断
- 设备信息
- 重启系统

服务状态

✖ 服务已停止

配置参数

设备编码

733e66a60ccec4aecb9d10f1da5e0192

可信设备ID

733e66a60ccec4aecb9d10f1da5e0192@box.netbooster

连接密码

.....

保存并应用

- 系统概览
- 网络配置
- 定向加速
- 智能组网
- API网关
- 网络诊断
- 设备信息
- 重启系统

重启系统

重启网关设备将导致当前的网络连接短暂中断（预计需要 1 分钟）。
请确认是否继续操作？

立即重启

可信大模型分控策略网关

边缘EDGE MLMtoken系列

简称：tokenBoss盒子

联系人：陈珂 博士 13775099862

常州万联云网信息技术有限公司

官网：www.macrounion.com