

ZENOVA

企业 AI 产品文档

配置 · 使用 · 结果核对

# Zenova 管理控制台操作指南

配置组织、模型和工具，验证权限与流量规则，找到运行问题的处理入口。

操作与使用　权限与边界　结果核对

文档版本

2026.09

核对日期

2026 年 9 月 16 日

# 目录

1. 当前导航与职责

2. 从旧界面迁移到当前入口

3. 建立组织与授权范围

4. 维护模型供应商

5. 发布模型服务与分配额度

6. 接入 MCP、技能与外部智能体

7. 流量规则、全局限额与网关扩展

8. 护栏、威胁情报与缓存

9. 用“运行总览”排查问题

10. 告警、导出与日常维护

11. 首次上线验收

> 2026.09 · 适用于 Web 管理控制台；部分能力由产品许可、平台身份和职责权限决定。

> 页面名称按当前源码核对；角色、许可和部署决定实际可用范围。正式操作前，以所在环境显示的配置与状态为准。

# 1. 当前导航与职责

登录后从管理入口进入控制台。菜单是按照当前身份与许可生成的，不是所有管理员看到同一套页面。

| 身份或职责 | 主要工作 | 不应混用的边界 |
| --- | --- | --- |
| 普通成员 | 对话、使用模型和工具、查看允许的资料 | 登录不代表有管理权 |
| 租户属主及有权管理人员 | 管理本组织配置、成员和角色 | 不自动拥有整套部署的运维权限 |
| 系统管理员 | 当前授权范围内的配置管理 | 与安全、审计、运营职责区分 |
| 安全管理员 | 当前授权范围内的凭据与安全策略 | 不自动获得角色编制权 |
| 审计员 | 当前授权范围内的审计与复核 | 不应使用配置管理身份替代审计身份 |
| 运营 | 当前授权范围内的运行与处置 | 系统部署级操作仍受平台身份限制 |
| 平台 Root 及平台运维职责 | 管理平台层面的组织生命周期、部署及运维 | 不用于普通成员的聊天和工作区体验 |

“角色管理”维护当前层级的职责成员。平台层和租户层的角色编制分别管理；如果页面提示职责冲突，应按组织分工处理，不通过重复赋权绕开限制。




*说明图依据当前导航与文案整理。真实菜单还会按账号权限、许可与运行形态裁剪。*

# 2. 从旧界面迁移到当前入口

| 旧材料中的找法 | 当前找法 |
| --- | --- |
| 单独寻找模型用量、会话任务、调用记录 | 调用与会话进入“运行总览”的“请求”“会话”页签；用量与费用在“首页 → 用量” |
| 直接寻找独立流量限额菜单 | “流量规则 → 全局限额” |
| 把规则试调当作第三个常驻页签 | 在规则清单使用“验证规则”，进入独立验证页面 |
| 在模型服务页混找供应商管理 | 业务服务在“模型服务”；上游维护在“模型供应商” |
| 认为每个管理员都能更新、备份 | “系统更新”“备份与恢复”为平台范围入口 |
| 按旧截图找审批中心 | 当前默认隐藏相关菜单；需要时核对许可并由管理员提供适用入口 |
| 在 Web 平台找云端连接、桌面宠物 | 属于桌面端专用个人设置，不是 SaaS Web 的常驻菜单 |

当前同样默认隐藏的入口包括管理台提示词库、API 服务、知识库管理、网关实例和网关运行参数等。隐藏不等于接口不存在；不能据此承诺所有客户默认可见或可用。

网关域名接入不在侧栏中：从“运行总览 → 会话”打开会话详情，在网关提示条点“查看接入方式”进入。平台运行日志同样不在侧栏中，只在平台控制台提供：从“网关与网络 → 网关 → 节点”进入节点详情，点“查看节点日志”。

# 3. 建立组织与授权范围

1. 进入“成员与部门”，按组织结构维护部门和成员，核对成员状态与归属。
2. 进入“职能组”，为跨部门职责建立需要的成员集合。
3. 有角色编制权限时进入“角色管理”，为系统、安全、审计和运营职责配置成员；提交后核对对应人员的实际菜单与操作权限。
4. Zenova 需要按团队或项目组织工作时，维护“工作空间”及其成员角色。
5. 在模型和能力详情中分配使用范围。组织归属、管理角色与资源调用权限分别检查。
6. 用普通账号验证可见和可调用，再用无权账号验证拒绝。

配置 SSO 时，从“SSO 配置”核对身份提供方、回调地址与用户映射。配置后应保留一个经组织批准的可恢复管理入口，并分别验证新用户与已有用户的归属。

“账号与会话安全”（“系统设置 → 全局设置”中的页签）用于相应账号与会话管理。撤销会话、调整登录策略等操作会影响使用者，应按组织流程确认对象并检查结果。

# 4. 维护模型供应商

模型供应商代表上游连接、认证及相关模型配置。模型服务是面向业务的可使用服务；二者不应混为一份清单。

1. 进入“搭建 → 模型供应商”，选择新增或维护现有供应商。
2. 按实际协议配置地址与认证。可接外部 SaaS 模型，也可接企业内网推理端点；不要只凭兼容名称推断所有能力相同。
3. 核对可用模型、价格或配额设置及页面内常驻展示的连通测试结果（多网关节点部署下，测试会按节点逐一探测并展示各节点的连通状态、耗时与失败层级；模型列表明确标示来自「供应商接口实时获取」还是「预置目录回退」；若部分节点暂不可达但需先行保存草稿，可勾选页面显示的「仍要保存」确认项）。
4. 在供应商详情页的「探活记录」中可按网关节点筛选历史探活结果，排查特定节点的网络、代理或证书问题。
5. 使用最小请求验证网络、TLS、认证、模型名称及响应格式。
6. 对实际业务所需的流式、图像、工具调用和结构化输出追加样例。

| 异常 | 排查方向 |
| --- | --- |
| 无法连接 | 地址、DNS、出站策略、代理、TLS 和上游状态 |
| 认证失败 | 凭据类型、有效期、上游账号与授权范围 |
| 找不到模型 | 上游模型标识、启用状态、目录刷新与权限 |
| 普通文本可用但工具失败 | 所选模型和接口是否支持实际工具调用格式 |
| 价格或计费信息不完整 | 价格配置是否发布、归属范围与币种是否正确 |

私有部署接公网模型时，提示与业务内容可能外发。将此上游纳入数据流向清单，再决定允许哪些业务使用。

# 5. 发布模型服务与分配额度

## 创建并验证服务

1. 进入“搭建 → 模型服务”，创建或编辑业务需要的服务。
2. 选择已准备好的模型来源，填写名称和说明，明确适用任务与数据边界。
3. 保存后进入详情，按基本信息、调试、可用范围、用量和审计等实际页签检查。
4. 先完成调试，再配置目标租户、工作空间、部门、职能组或成员等可用对象。可选对象以当前产品与权限为准。
5. 在适用范围设置额度或限制，核对单位、周期和超限处理。
6. 用目标普通成员进入模型广场，获取正确归属的 API 密钥，发起受控调用。

模型服务列表当前有“模型服务”和“API 密钥管理”两个顶层页签。模型供应商列表有“模型供应商”“模型&定价”和“分类管理”三个顶层页签。不要把仍有路由的全部模型或应用管理页面写成常驻页签；详情页面和列表页面功能不同。




*2026-09-16 隔离测试环境截图。列表是 Mock 测试服务，金额与用量不是对外报价或性能证据。*

### 凭据与计费归属

创建 API 密钥时需要有效的服务与计费归属；多候选项应明确选择。候选配置变更或正在下发时，刷新并重新确认，避免按过期价格或范围创建。

待审批状态与已签发状态分别处理。既有密钥的明文复制受当前权限与接口支持限制，不把“一次性展示”作为全产品统一承诺。

### 额度不是一个数字

| 控制 | 管什么 | 应如何验收 |
| --- | --- | --- |
| 使用授权 | 哪个主体能使用哪个服务 | 有权成功，无权拒绝 |
| 服务或主体额度 | 当前支持范围内的消耗限制 | 核对周期、单位与归属，触达边界后验证结果 |
| 全局限额与规则内限额 | 适用请求的次数、并发等流量条件 | 正例、超限与恢复场景分别验证 |
| 预算提醒 | 提示接近或超过预算 | 验证通知与记录，不把提醒当成硬阻断 |

## 6. 接入 MCP、技能与外部智能体

“MCP 服务”维护工具服务的连接、工具目录和开放范围。服务可来自内网，也可来自外部 SaaS；后者还需外部账号、网络与服务许可。

1. 添加或导入服务，按页面要求配置传输、地址与认证。
2. 发现或刷新工具，核对工具名称、说明、参数及读写性质。测试连接与工具发现会在全部在线网关节点上执行，并在当前页面直接展示逐节点测试结果与失败原因。
3. 配置共享凭据或要求使用者完成个人授权，避免把个人凭据当作全组织共享账号。
4. 设置发布状态与使用范围，并检查数据等级；MCP 服务的工具级访问控制、参数匹配规则和请求体积上限统一在“路由策略 → 流量规则”中配置。
5. 用普通成员复制连接信息，再用只读样例验证客户端到网关、网关到上游的完整路径。
6. 变更工具定义或权限后重新验证；停用前检查使用者与正在执行的任务。

“技能”“插件”和“外部智能体”分别维护不同的能力对象。技能可能依赖工具，插件可能包含技能和连接配置，外部智能体有自己的服务接口。安装完成不自动证明依赖已经授权或可调用。

外部注册中心用于管理配置允许的能力来源。导入后仍需检查内容、依赖和权限；不能把来源存在当作可直接对全体成员开放。

# 7. 流量规则、全局限额与网关扩展

当前入口为“路由策略 → 流量规则”。页面分为“规则清单”和“全局限额”，并提供“验证规则”。

1. 先明确规则保护对象与调用类型，准备应命中和不应命中的请求样例。
2. 在规则清单创建模型路由或增强规则，按页面设置条件、动作和顺序。
3. 路由目标使用已准备好的可用服务；调用方身份使用对应入口实际提供的可信信息，不能把客户端任意 Header 当作可信身份。
4. 设置适用的限额或全局限额。次数、并发、超时和熔断保护不同对象，不互相替代。
5. 保存后查看下发状态；用“验证规则”核对命中链和目标，再发起真实受控请求。
6. 在“运行总览 → 规则命中”页签和“运行总览 → 请求”核对结果，保留请求标识及发生时间。

熔断会根据失败或慢调用暂时拒绝后续请求；故障转移负责选择其他目标。熔断本身不等于自动切换备用模型。模拟也不能代替真实熔断、排队或上游故障测试。

“网关扩展”位于路由策略分组，与“插件”不同。按插件说明核对包、配置和绑定范围，再通过实际调用验证执行效果。

字段定义、条件组合、分流、限额和异常处置详见配套《路由策略用户手册》。维护规则顺序、启停和插件绑定后，重新验证受影响的业务样例。

# 8. 护栏、威胁情报与缓存

## 内容与访问控制

“护栏”用于配置词库、必答、敏感信息规则与内部系统名单等共享资产；具体的输入/输出检测、图文联合检测、多轮拼接诱导防护及工具调用防护统一挂在“路由策略 → 流量规则”的规则动作上（原 MCP 服务详情页的独立护栏页签已并入路由策略）。按业务确定输入与输出方向、检测项、作用范围及放行、记录、脱敏或阻断等实际动作。

先用合成样例验证应拦、应放和边界情况，再扩大生效范围。不同检测器、入口和异常处理有自己的行为，不把“无法判定一律拒绝”写成所有功能的默认状态。

“威胁情报”是“护栏 › 护栏规则”的一个页签：租户在此接入自己的恶意库（公开源 URLhaus、ThreatFox、Feodo，或自定义 HTTP 批量源，每个租户最多 8 个，各自独立启用、设置同步间隔与凭据），查看各恶意库的同步与生效状态，并维护例外名单。平台不提供默认恶意库，平台管理员没有威胁情报入口。是否检查、命中后拦截还

是仅记录，在流量规则中添加“恶意来源”动作配置，不开启即不检查；检查覆盖网关代理出口、工具执行出口与影子流量。从早期版本升级后，需要重新接入恶意库并在流量规则中开启检查，详见《Zenova 用户手册》第 25.16 节。

边缘 WAF（Web 应用防火墙）处理入口层面的适用流量。当前交付包含边缘 WAF 集成，实际启用模式与规则由部署配置决定，不能把组件存在当作所有请求正文已经检测。

### 语义缓存

平台层的“语义缓存”管理所需服务与相关状态；业务规则决定何时适用。先核对当前部署包含必要组件，再选择适合缓存的请求。

1. 使用低风险、允许复用结果的样例评估命中质量。
2. 按当前可用的观察或生效模式设置，并确认租户、调用方、模型与数据范围。
3. 对重复问题、不同身份、不同模型和内容变更分别验证。
4. 命中后核对上游调用、延迟和账单；失效或超时行为按实际返回验证。

涉及实时数据、身份敏感内容或必须每次重新执行的工具，不应未经验证就复用缓存。缓存收益以实际测量为准。

## 9. 用“运行总览”排查问题

| 页签或入口 | 解决的问题 |
| --- | --- |
| 运行总览 → 请求 | 这次调用是否完成、去了哪里、返回什么结果 |
| 运行总览 → 会话 | 一组会话或任务的运行情况 |
| 运行总览 → 规则命中 | 哪条流量规则匹配并采取动作 |
| 运行总览 → 告警 | 当前触发的告警及其关联事实 |
| 首页 → 用量 | 当前统计范围用了多少 Token、调用次数与费用 |
| 操作审计 | 谁在何时修改、创建或删除了什么配置 |
| 平台运行日志（仅平台控制台：节点详情 → “查看节点日志”） | 进一步查询技术错误与运行日志 |

用量和成本内容受相应许可与权限影响。没有某个页签时先核对授权，不直接认定采集失败。

排查顺序：先确定时间和时区，再确定租户或工作空间，找到请求编号或对象，查看结果和原因，然后进入规则、供应商或工具配置处理。修正后用同类最小请求重新验证。

网关调用记录、业务观测和操作审计用途不同。外部应用内部步骤只有在相应观测数据接入后才能展示；不能依据网关一跳自动还原所有业务执行。

# 10. 告警、导出与日常维护

“告警规则”配置当前层级可管理的规则与通知。平台身份与租户身份看到的范围不同。先测试通知连接（测试结果与失败分类直接显示在当前配置卡片/对话框内，不会一闪而过），再验证规则触发、恢复和接收人；空白列表可能是当前范围没有规则。LDAP、OIDC、IM 渠道、镜像源、威胁情报恶意库以及网关出口代理的连通测试均在页面内常驻展示测试结果（出口代理测试会逐节点展示各网关节点经该代理的连通耗时与状态）。

删除租户、重置用户密码、替换或删除网关证书与域名、修改需重启的网关运行参数、启用旧版 MCP 身份以及退役网关节点等高风险操作，均通过统一的二次确认对话框执行（对话框写明操作动作与目标资源名称；删除租户与退役网关节点还需输入资源名称确认）。

“报表导出”用于当前支持的导出任务。核对导出范围、时间窗、权限与生成结果，下载后检查数据完整性；不能把导出成功当作完成数据库恢复备份。

平台范围的“系统更新”“备份与恢复”“系统运行监控”“部署配置”“SSRF 出站防护”和“网关入口”等入口由平台运维人员操作。SaaS 租户按服务约定提交请求，不自行照私有部署命令管理托管基础设施。

私有部署变更前记录版本与配置，明确维护窗口，完成适用备份与恢复准备；操作后核对正常调用、权限拒绝、记录、通知与备份状态。不要根据旧版文档推断升级必然无中断或不需要数据迁移。

# 11. 首次上线验收

| 对象 | 正常路径 | 拒绝或失败路径 |
| --- | --- | --- |
| 组织与角色 | 指定人员能完成职责操作 | 无权人员看不到或无法执行该操作 |
| 模型服务 | 普通成员可获取凭据并调用 | 无权、无效凭据和额度边界按预期处理 |
| MCP 工具 | 授权完成且只读测试成功 | 缺授权、上游失效与策略拒绝可识别 |
| 流量规则 | 正例命中正确目标 | 反例、超限与恢复行为可核对 |
| 护栏 | 正常内容通过 | 合成敏感内容执行指定动作 |
| 运行记录 | 请求、主体、结果和用量可关联 | 缺失或延迟有明确排查方法 |
| 运维 | 通知、备份和职责交接完成 | 实际拓扑中的故障恢复经过演练 |

把验收记录交给运维及业务负责人，包含版本、环境、请求标识、预期与实际结果，以及仍未完成的项目。未验收能力不标为已上线。