

# 辰极智脑 - 系统管理需求说明书

## 一、修订记录

修订日期	修订内容	修订人员	审核人员
2025-12-20	辰极智脑系统管理功能初稿	辰极智脑团队	辰极智脑团队

## 二、需求背景与目标

### 2.1 编写目的

本文档旨在明确"辰极智脑"系统中"系统管理"模块的功能需求。该模块作为系统管理员的核心控制台，负责全局配置管理、资源监控、性能调优及维护操作，是保障系统稳定运行的基石。

### 2.2 产品背景

"辰极智脑"是一个复杂的 AI 中台系统，包含知识检索引擎、知识库、大模型接入等多个子系统。为了降低运维复杂度，系统提供了一个统一的管理界面，实现：

- 集中管控**：一站式管理系统参数、摇光组件配置及资源配额。
- 可观测性**：实时监控后台任务状态、系统健康度及关键性能指标。
- 主动维护**：提供性能基准测试、知识库清空重建等高级维护工具。

## 三、用户角色与权限

角色	职责描述	权限范围
平台管理员	负责系统底层的稳定性和全局配置。	完全权限。拥有所有配置的修改权及破坏性操作权限。
企业管理员	负责本企业业务的运行情况。	无权限/受限访问。通常不可见此页面。

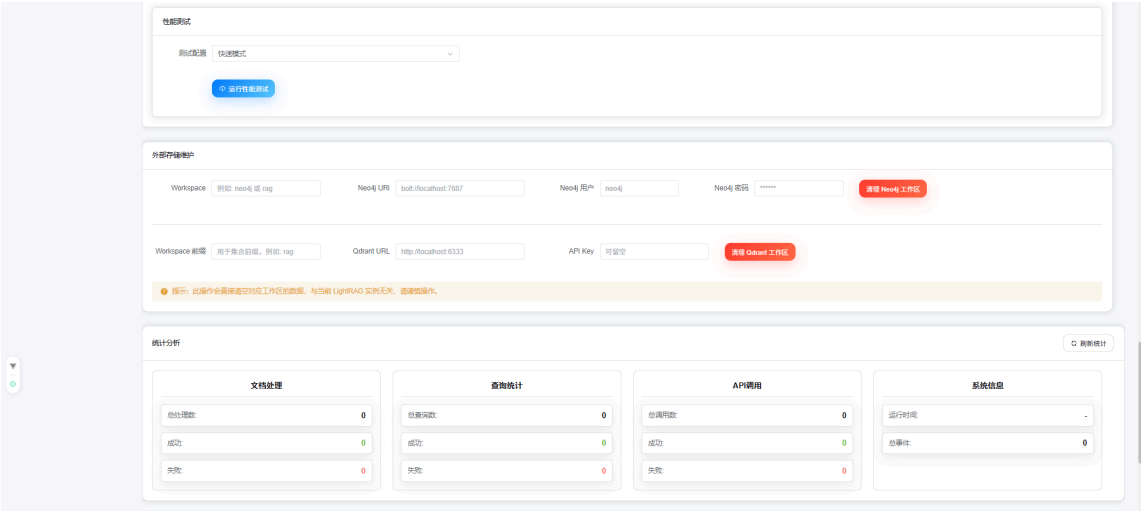
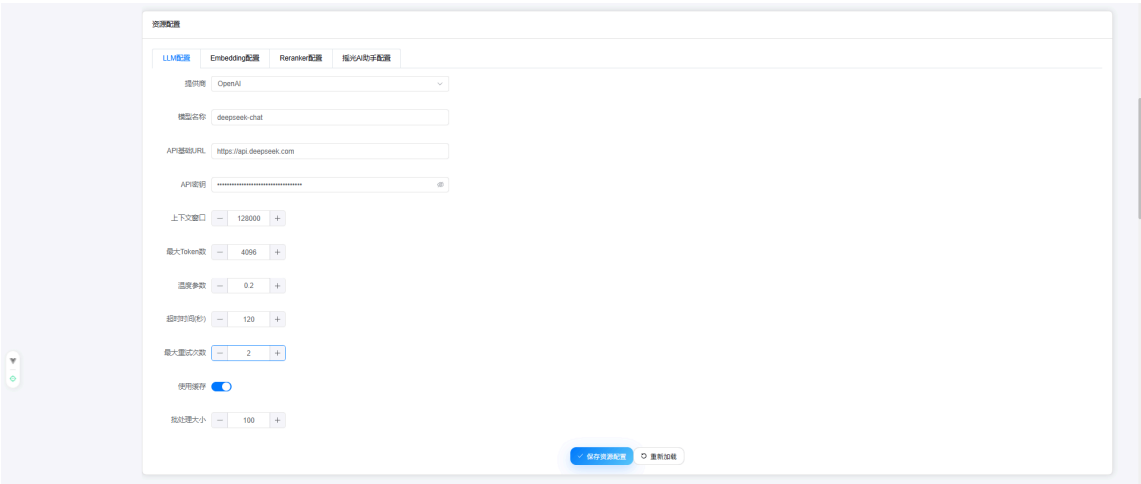
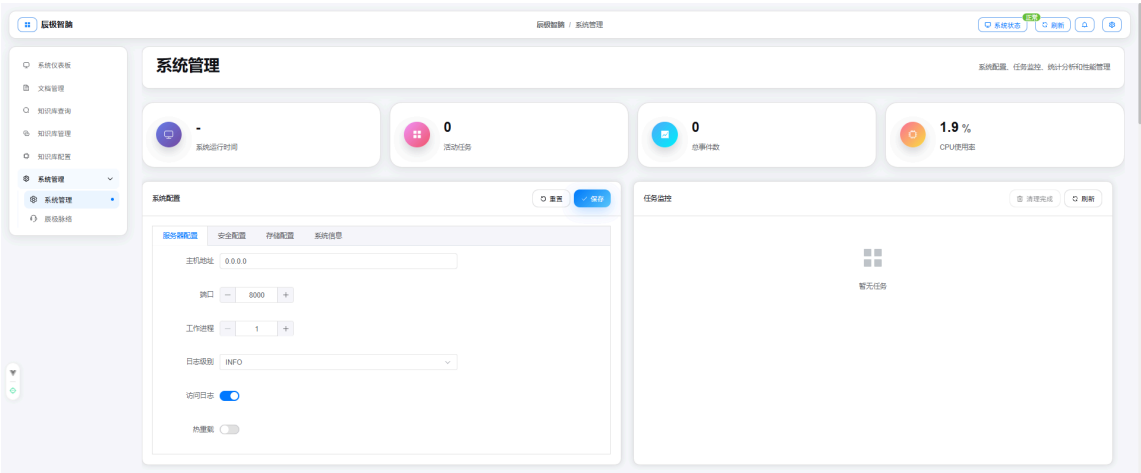
## 四、功能架构

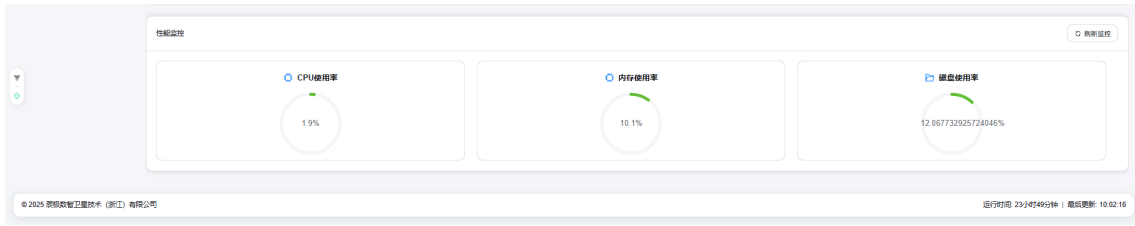
系统管理模块采用模块化卡片布局，主要包含以下功能区块：

- 概览仪表盘**：系统运行时长、总事件数等宏观指标。
- 系统配置**：基础参数设置。
- 任务监控**：后台任务的实时追踪与干预。
- 资源配置**：知识检索引擎参数及摇光组件配置。
- 性能配置**：性能模式切换与基准测试。
- 统计与监控**：详细的图表化监控数据。

## 五、功能性需求

原型图





## 5.1 系统概览

### 功能描述

在页面顶部展示系统关键状态摘要：

- **运行时长：**系统自上次启动以来的持续运行时间，以"X天X小时X分钟"格式展示。
- **总事件数：**系统处理的累计请求或事件数量。
- **活跃任务数：**当前正在执行的后台任务数量。
- **资源状态：**以红绿灯形式展示 CPU 和内存的简要负载状态。

## 5.2 系统配置

### 功能描述

提供系统级全局参数的查看与修改。

- **配置项：**
  - **日志级别：**可选 调试 / 信息 / 警告 / 错误。
  - **缓存策略：**缓存过期时间、最大条目数。
  - **请求限制：**全局每秒请求阈值。
- **操作：**
  - **保存：**提交修改，系统无需重启即可生效。
  - **重置：**恢复至系统默认配置。

## 5.3 任务监控

### 功能描述

实时追踪系统中运行的后台任务（如知识库构建、批量文档解析等）。

- **任务列表：**展示任务编号、类型、状态（排队中/进行中/已完成/失败）、进度条。
- **干预操作：**
  - **刷新：**手动同步最新任务列表。
  - **取消：**强行终止正在运行的任务。
  - **重试：**对失败的任务发起重新执行。
  - **清除历史：**一键移除所有已完成或已失败的历史记录。

## 5.4 资源与摇光配置

### 功能描述

管理底层计算资源及"摇光"组件的连接参数。

- **配置锁定：**提供"锁定配置"开关。开启后，所有参数变为只读状态，防止误操作。

- 知识检索参数：检索召回数量、文本分块大小、分块重叠字符数等核心参数。
- 摇光配置：
  - 预设方案：提供多种预置方案供快速切换（如"高性能模式"、"省资源模式"）。
  - 连接参数：摇光服务的服务地址、访问密钥等，密钥需自动脱敏显示。
- 知识库维护：
  - 状态统计：展示当前知识库中的向量总数、文档总数。
  - 清空知识库：高危操作按钮。
    - 交互要求：点击后必须弹出红色警示框，需用户明确二次确认后方可执行清空操作。

## 5.5 性能配置与测试

### 功能描述

- 模式管理：
  - 预置模式：提供多种系统运行模式供快速切换（如"均衡模式"、"吞吐优先"、"延迟优先"）。
  - 自定义模式：支持新建、编辑、删除自定义的参数组合。
- 基准测试：
  - 功能：一键发起性能测试，评估当前配置下的系统表现。
  - 参数：可设置并发数、测试时长。
  - 结果：测试完成后展示每秒处理量、平均延迟、99分位延迟等指标。

## 5.6 统计与性能监控

### 功能描述

- 业务统计：各企业的请求量分布、各模型的使用占比。
- 性能图表：以折线图展示近 1 小时的请求量趋势、响应时间波动。
- 存储维护：展示磁盘使用率、对象存储的存储配额，提供过期文件清理工具。

---

# 六、非功能性需求

## 6.1 响应性

- 加载状态：各卡片数据加载时需展示骨架屏或加载动画。
- 实时性：任务列表和监控图表应支持自动刷新，延迟不超过 5 秒。

## 6.2 健壮性

- 异常处理：任何配置保存失败均需弹出明确的错误提示，禁止静默失败。
- 防误触：所有破坏性操作（删除任务、清空库、重置配置）必须有二次确认机制。