

OctopusAI - AI应用平台 使用说明



目录

- 平台概述
- 界面介绍
- AI应用管理
- AI应用设计
- 工作流设计
- GPU管理
- 任务执行

平台概述

OctopusAI 是一款通用的AI应用开发和执行平台，提供了从应用设计、工作流编排到GPU资源管理的完整解决方案。平台支持动态加载各类AI应用，通过可视化的工作流设计器编排复杂的AI处理流程，并自动调度GPU算力资源执行任务。

核心功能

- **AI应用设计**: 创建和配置各类AI应用，支持动态表单生成
- **工作流设计**: 可视化拖拽式设计复杂的AI处理流程
- **GPU实例管理**: 集成GPU租用服务，支持实例的创建、启停和管理
- **远程任务执行**: 通过SSH安全连接远程GPU主机执行AI任务
- **实时进度监控**: 可视化展示任务执行流程和进度状态
- **历史记录管理**: 自动保存任务记录，支持结果查看和配置复用

- **批量处理**: 支持设置执行次数，自动进行批量任务处理

平台特点

- **应用独立性**: 每个AI应用独立配置，互不干扰
- **配置持久化**: 自动保存应用配置和执行历史
- **资源优化**: 智能调度GPU资源，支持任务完成后自动关机
- **网络加速**: 支持学术加速和SSH境外代理，适配不同网络环境
- **扩展性强**: 支持通过配置文件添加新的AI应用

应用场景

平台可应用于多种AI场景，例如：

- 视频生成与编辑（如商品讲解视频生成）
- 图像处理与生成
- 语音合成与识别
- 自然语言处理
- 多模态AI任务
- 自定义AI工作流

界面介绍

平台采用三栏布局设计，功能分区清晰，操作便捷直观。

整体布局

区域	功能说明
顶部导航栏	包含应用切换、功能入口和全局操作按钮
左侧面板	应用配置区域，根据当前应用动态生成配置表单
中间面板	历史记录区域，展示应用执行历史和结果
右侧面板	执行流程区域，实时显示任务执行步骤和状态

顶部导航栏

标签页

- **应用设计**: 当前激活标签，显示选定应用的配置界面
- **AI应用设计**: 链接到AI应用设计器，用于创建新应用
- **AI工作流设计**: 链接到工作流设计器，用于设计处理流程

操作按钮

- **刷新**: 刷新当前页面数据，重新加载应用状态
- **AI应用设计** (链接) : 快速跳转到应用设计器
- **AI工作流设计** (链接) : 快速跳转到工作流设计器

左侧配置面板

左侧面板内容根据当前选择的应用动态变化，通常包含以下区域：

应用介绍区

- **应用来源**: 显示应用提供者信息
- **联系作者**: 获取应用技术支持

- **项目主页**: 访问应用项目主页
- **编辑配置**: 编辑应用配置文件

参数配置区

根据应用需求动态生成的表单字段，常见类型包括：

- **文本输入框**: 输入产品描述、提示词等文本内容
- **文件选择器**: 上传音频、图片、视频等参考文件
- **多文件列表**: 管理多个参考文件，支持添加和删除
- **下拉选择框**: 选择预设选项
- **复选框**: 启用/禁用特定功能选项
- **数字输入**: 设置数量、次数等数值参数

提示: 带 * 标记的字段为必填项，必须填写才能执行任务。

算力中心区

- **主机选择**: 下拉列表选择可用的GPU实例
- **状态显示**: 显示实例运行状态（运行中/已停止等）
- **SSH连接信息**: 主机地址、端口、用户名、密码
- **快捷操作**: 复制密码、SSH连接、主机日志等

执行控制区

- **处理选项**: 初始化安装、学术加速、SSH代理等复选项
- **执行次数**: 设置批量执行的次数
- **停止任务**: 中断当前执行的任务

中间历史记录面板

展示当前应用的所有执行历史记录，每条记录包括：

- **任务标题**: 显示应用名称和执行信息

- **执行时间:** 任务启动的日期和时间
- **配置参数:** 显示关键配置信息
- **生成结果:** 输出文件的保存路径

每条记录提供"查看"、"使用"、"删除"操作按钮。

右侧执行流程面板

实时显示当前任务的执行步骤，采用层级结构展示：

- **主步骤:** 显示任务的主要阶段（如步骤1、步骤2...）
- **子步骤:** 显示每个主步骤的详细操作
- **状态指示:** 用颜色标识步骤的执行状态
 - 绿色 - 已完成
 - 红色 - 执行中
 - 灰色 - 待执行

AI应用管理

平台支持管理多个AI应用，每个应用独立配置，互不影响。

应用类型

平台支持两种类型的应用：

- **单场景应用:** 执行单一特定任务的应用（如"商品讲解视频生成-单场景"）
- **多场景应用:** 支持多个场景或模式的应用

应用切换

通过顶部导航栏的"AI应用设计"链接进入应用管理界面：

- 查看所有已安装的应用列表
- 选择要使用的应用并加载其配置界面
- 创建新的AI应用
- 编辑现有应用的配置

应用配置

每个应用的配置界面包含以下步骤：

步骤1：填写应用参数

根据应用需求填写相应的参数字段：

- **文本类参数**: 在输入框中填写内容
- **文件类参数**: 点击"浏览"或"添加文件"上传文件
- **选项类参数**: 从下拉列表选择或勾选复选框

步骤2：选择GPU实例

在"算力中心"区域选择执行任务的GPU主机：

- 从下拉列表选择可用的GPU实例
- 确认实例状态为"运行中"
- 如需新建实例，点击"租用GPU主机"

步骤3：配置执行选项

根据需要设置执行控制选项：

- 勾选"初始化安装" (首次使用该GPU时)
- 根据服务器位置选择"学术加速"或"SSH境外代理"
- 设置"执行次数" (批量处理时)
- 勾选"完成后关机"以节省费用

步骤4：启动任务

配置完成后，任务将自动开始执行：

- 观察右侧"执行流程"面板查看实时进度
- 需要中断时点击"停止任务"按钮

步骤5：查看结果

任务完成后，在"历史记录"面板查看结果：

- 点击"查看"打开输出文件夹
- 点击"使用"可复用本次配置

应用示例：商品讲解视频生成

以下是平台中一个典型应用的配置示例：

参数名称	说明
产品说明 *	输入产品名称或描述，用于生成营销文案和视频内容
音色参考文件 *	上传10秒以内的音频文件，用于提取说话人音色
参考图像 *	上传产品相关的原始图片和参考图片，支持多图
首图生成提示词 *	描述对视频首图的编辑要求和期望效果

注意：不同应用的参数配置各不相同，请根据具体应用的要求填写相应的参数。

AI应用设计

平台提供AI应用模块生成器，支持通过开源项目URL自动生成应用配置文件。用户无需编写代码，即可快速添加新的AI应用。

应用设计器访问

- 点击顶部导航栏的“**AI应用设计**”链接
- 或在应用配置界面点击“**编辑配置**”进入编辑模式

新建应用

通过项目URL自动创建新应用配置：

步骤1：输入项目信息

在应用设计器中：

- 在“项目网址”输入框粘贴项目URL
- 支持GitHub项目链接（如：<https://github.com/xxx/xxx>）
- 支持ModelScope项目链接（如：<https://www.modelscope.cn/models/xxx/xxx>）
- 在“生成提示”框输入额外要求（可选）
- 勾选“使用Claude Code”可启用交互式编辑（可选）

步骤2：生成配置

点击“立即构建”按钮：

- 系统自动获取项目的README文档
- AI分析项目内容并生成应用配置
- 在“生成日志”区域显示实时进度

步骤3：预览配置

生成完成后：

- 在“生成结果预览”区域查看配置内容
- 检查应用名称、描述、参数等是否正确
- 如需修改，可调整“生成提示”后重新生成

步骤4：保存应用

点击“保存配置”按钮：

- 确认文件名（通常为app_id.json格式）
- 选择保存位置（默认为 `./task_configs/` 目录）
- 保存后应用自动出现在应用列表中

编辑应用

修改已有应用的配置：

步骤1：选择应用

在应用设计器中：

- 从“应用选择”下拉列表中选择要编辑的应用
- 当前配置自动加载并显示

步骤2：描述修改需求

在“生成提示”框中输入修改说明：

- 例如：“添加语言选择参数”、“修改默认值”、“修复错误”等
- 清晰描述可获得更好的修改效果

步骤3：生成修改

点击“立即构建”按钮：

- AI根据要求自动修改配置文件
- 在“生成结果预览”区域查看修改后的配置

步骤4：保存修改

点击“保存配置”覆盖原文件：

- 修改后的配置立即生效
- 刷新应用列表可看到更新

使用技巧

新建应用建议

- **选择项目**: 优先选择有详细README文档的项目
- **明确需求**: 在生成提示中清晰描述所需功能和参数
- **分步完善**: 先生成基础配置，再逐步添加高级参数
- **测试验证**: 保存后先运行简单任务测试配置是否正确

编辑应用建议

- **备份配置**: 修改前先备份原配置文件
- **小步修改**: 每次只修改一个或少数几个参数
- **验证效果**: 每次修改后测试应用是否正常工作

注意事项

- 应用ID (app_id) 必须唯一，不能与已有应用重复
- 应用名称建议使用"中文功能-项目名"格式
- 生成完成后建议仔细检查配置文件的完整性
- 复杂的应用配置可能需要多次调整才能达到理想效果

工作流设计

平台提供可视化工作流设计器，支持通过拖拽方式设计复杂的AI处理流程。工作流由多个AI应用组件组合而成，实现端到端的自动化任务处理。

访问工作流设计器

- 点击顶部导航栏的"AI工作流设计"链接
- 或在应用配置界面点击"编辑配置"进入流程编辑

设计器界面

工作流设计器采用专业的可视化设计界面，支持拖拽式操作和实时预览。

界面分区说明

区域	功能说明
顶部工具栏	包含工作流选择下拉框、文件操作按钮（新建/打开/保存）、自动排列、验证工作流等快捷操作
左侧应用库	列出所有可用的AI应用组件，支持搜索和快速添加。每个应用显示名称和简短描述
中间画布	大型工作区域（3000×3000像素），支持滚动和缩放。可自由拖拽节点、创建连接
右侧属性面板	包含三个选项卡： <ul style="list-style-type: none">属性：配置工作流元信息、添加用户输入节点预览：查看生成的JSON配置文件AI生成：使用AI自动生成工作流配置

画布元素

- 应用节点**：方形卡片，显示应用名称和连接状态。选中时显示高亮边框
- 用户输入节点**：圆形节点，代表需要用户提供的输入参数
- 连接线**：带箭头的连线，表示数据流转方向。连接线通过配置输入项后自动绘制
- 网格背景**：帮助对齐节点，提升视觉清晰度

节点状态说明

- 普通状态**：白色背景，黑色边框
- 选中状态**：橙色边框高亮显示
- 已配置输入**：节点左侧显示绿色指示点

- **未配置输入**: 节点左侧显示红色指示点

创建工作流

步骤1：添加应用节点

从左侧应用列表添加组件到画布：

- 在左侧面板浏览或搜索应用
- 点击应用名称或“添加”按钮
- 节点自动添加到画布中心
- 拖拽节点调整位置

步骤2：配置节点输入

为节点的输入参数配置数据来源。这是工作流设计最核心的操作。

打开配置对话框：

- **方法1**: 右键点击节点，选择“配置输入项...”
- **方法2**: 选中节点后按 **F2** 键

输入配置对话框说明：

- 对话框标题显示节点名称
- 列出该节点的所有输入参数
- 每个参数显示：参数标签、是否必填、当前配置状态

5种输入类型详解：

输入类型	说明	使用场景
输入组件	使用右侧"工作流输入"中定义的用户输入组件	需要用户在执行时提供的参数（如文本、文件等）
模块预置输入	由用户在执行时通过输入框提供的值	临时性的、不需要作为工作流输入的参数
节点引用	使用其他应用节点的输出结果	节点间的数据传递，形成工作流
默认值	使用应用配置中预设的默认值	固定参数，不需要用户修改
空值	不提供任何值	可选参数，留空不使用

自动配置规则：

- **必填参数 + 有默认值** → 自动使用"默认值"
- **必填参数 + 无默认值** → 自动使用"模块预置输入"
- **可选参数** → 自动使用"空值"

配置完成后：

- 点击"保存"按钮应用配置
- 系统自动绘制连接线
- 节点左侧显示绿色/红色指示点表示配置状态

步骤3：添加工作流输入

在右侧"属性"选项卡中：

- 点击"工作流输入"区域的" + 添加用户输入 "
- 配置输入参数：
- **参数ID**: 唯一标识符 (如: input_text)
- **参数标签**: 显示名称 (如: 输入文本)

- **参数类型**: 文本、文件、数字等
- **是否必填**: 勾选表示必须提供
- 添加后可在配置节点输入时选择

步骤4：预览和验证

切换到“预览”选项卡：

- 查看生成的JSON配置文件
- 检查节点连接是否正确
- 验证所有必填参数是否已配置

点击“验证工作流”按钮：

- 系统自动检查配置完整性
- 提示未配置的必填参数
- 检查循环依赖等逻辑错误

步骤5：保存工作流

点击工具栏的“保存”按钮：

- 输入工作流文件名 (如: my_workflow.json)
- 选择保存位置 (默认为 `./task_configs/` 目录)
- 保存后可在应用列表中找到该工作流

GPU管理

平台集成了GPU租用服务，支持弹性GPU资源管理，按需使用，成本可控。

租用GPU实例

步骤1：选择GPU型号

在“算力中心”区域：

- 点击“租用GPU主机”按钮
- 从列表中选择合适的GPU型号：
 - RTX 3090：适合中小型模型训练和推理
 - RTX 4090：高性能计算，适合大型模型
 - A100：企业级GPU，超大模型训练
 - H100：最新一代，极致性能

步骤2：配置实例参数

- **区域选择**：选择服务器地理位置（影响网络延迟）
- **计费方式**：
 - **按小时计费**：适合短期任务
 - **按天计费**：适合长期项目
- **磁盘空间**：根据需求选择存储大小
- **镜像选择**：选择预装环境的系统镜像

步骤3：启动实例

- 点击“立即租用”按钮
- 系统自动创建并启动GPU实例
- 等待实例状态变为“运行中”
- 获取SSH连接信息

实例状态监控

在“算力中心”区域可以查看：

- **实例名称**：GPU实例的标识

- **运行状态**: 运行中、已停止、已创建等
- **计费状态**: 正在计费、已停止计费
- **运行时长**: 实例已运行的时间
- **预计费用**: 当前费用预估

SSH连接配置

连接信息

- **主机地址**: GPU服务器的IP地址
- **SSH端口**: 通常为22
- **用户名**: 登录用户名 (通常为root)
- **密码**: 自动生成的密码

快捷操作

- **复制密码**: 一键复制SSH密码
- **测试连接**: 验证SSH连接是否正常
- **打开终端**: 直接打开SSH终端
- **查看日志**: 查看GPU实例运行日志

成本优化建议

- **任务完成后关机**: 及时停止实例以节省费用
- **选择合适配置**: 根据任务需求选择GPU型号
- **利用spot实例**: 使用竞价实例降低成本 (可能被中断)
- **定期清理**: 删除不再使用的实例

任务执行

任务启动

配置完成后，任务将自动开始执行：

1. **初始化阶段**
2. SSH连接到GPU主机
3. 检查环境依赖
4. 安装必要的软件包（如勾选“初始化安装”）

5. 执行阶段

6. 上传输入文件到GPU服务器
7. 执行AI模型推理
8. 监控执行进度

9. 完成阶段

10. 下载输出文件到本地
11. 保存执行记录
12. 自动关机（如勾选“完成后关机”）

进度监控

在右侧“执行流程”面板实时查看：

- **主步骤**：显示任务的主要阶段
- 步骤1：初始化环境
- 步骤2：上传文件
- 步骤3：执行AI模型
- 步骤4：下载结果
- **子步骤**：每个主步骤的详细操作

- **执行状态:**
- 绿色 - 已完成
- 红色 - 执行中
- 灰色 - 待执行
- 红色叉 - 执行失败

任务中断

如需中断任务：

- 点击左侧面板的"停止任务"按钮
- 系统发送中断信号
- GPU服务器停止执行
- 已生成的结果会保存

结果查看

任务完成后：

1. **查看历史记录**
2. 在中间"历史记录"面板找到任务
3. 查看执行时间和配置信息
4. **查看输出文件**
5. 点击"查看"按钮
6. 系统打开输出文件夹
7. 查看生成的文件
8. **复用配置**
9. 点击"使用"按钮
10. 系统自动加载本次配置
11. 可直接执行或修改后执行

批量执行

如需处理多个任务：

1. **设置执行次数**
 2. 在"执行次数"输入框设置数量
 3. 系统会依次执行多次任务
 4. **自动队列**
 5. 任务自动加入执行队列
 6. 前一个任务完成后自动开始下一个
 7. 全部完成后显示汇总信息
 8. **查看批量结果**
 9. 历史记录显示所有执行记录
 10. 每条记录独立标识
 11. 可单独查看每个结果
-

常见问题

Q1：如何添加新的AI应用？

A：点击顶部导航栏的"AI应用设计"链接，输入项目URL（GitHub或ModelScope），系统会自动生成应用配置。

Q2：工作流中的节点如何连接？

A：右键点击节点选择"配置输入项"，选择"节点引用"作为输入类型，然后选择源节点的输出。系统会自动绘制连接线。

Q3：GPU实例如何计费？

A: 支持按小时和按天计费。任务完成后建议及时关机以节省费用。可在“算力中心”区域查看实时费用。

Q4：如何使用学术加速？

A: 在执行选项中勾选“学术加速”，平台会通过学术网络加速访问境外资源。

Q5：任务执行失败怎么办？

A: 在右侧“执行流程”面板查看错误步骤，点击红色叉号查看详细错误信息。常见问题包括：网络连接失败、依赖缺失、配置错误等。

Q6：如何导出工作流配置？

A: 在工作流设计器的“预览”选项卡中，可以查看和复制JSON配置文件，也可点击“保存”将配置保存到本地。

技术支持

如需帮助，请通过以下方式联系：

- **GitHub Issues**: 提交问题和建议
 - **项目主页**: 查看最新更新和文档
 - **联系作者**: 获取技术支持
-

OctopusAI - 让AI应用开发更简单！

最后更新时间: 2025-01-14