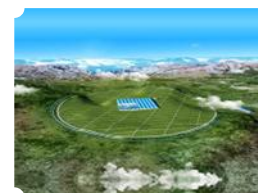




高能AI平台整体介绍



张正德 代表计算中心AI团队

2026.09.18



Content

1

背景

2

HAI平台简介

3

应用与服务

4

总结与展望

Part 1

背景



AI4S正在成为下一轮科技革命的制高点



主要经济体加速战略布局，科学智能已形成诺贝尔奖级突破



中国：国家战略部署

国务院“人工智能+”行动意见

“加快探索人工智能驱动的新型科研范式”

国家重点行动

2025年8月

推动科研平台与重大科技基础设施智能化升级



美国：争夺科技领导权

Genesis Mission 创世纪计划

整合超算、科学数据与AI
建设国家级科研平台

> 50亿美元

联邦支持承诺，2026年7月

首批278个入选项目
涉及342家参与机构



欧盟：布局科研AI前沿

AI in Science Strategy

建设RAISE平台
统筹欧洲科学AI资源

6亿欧元

科研算力支持安排，2025年10月

改善科研算力访问
计划将科学AI资助翻倍



科学：标志性突破

2024年诺贝尔化学奖

AlphaFold推动
蛋白质结构预测突破

约2亿种

蛋白质结构预测

190个国家、超过200万人使用
(诺奖官方2024年口径)

人工智能正在推动科研范式变革

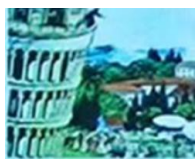


高能物理正处于**数据范式向智能范式转变的阶段**。

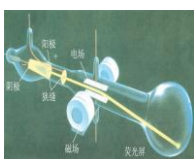
- 划分科研范式的主要依据之一是**科研工具的使用**，新工具的使用必然会提高科研效率，**催生新成果**；
- 科研范式演变的研究对象从**“单一现象”**到**“复杂系统”**；从研究**“确定”**到研究**“不确定”**。

实验科学：经验

描述记录自然现象，基于实验或经验的归纳为主；

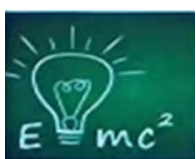


比萨斜塔 阴极射线

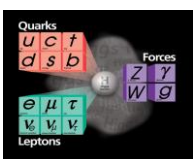


理论科学：理论

在自然现象基础上进行了抽象简化，构建数学模型；



相对论 标准模型

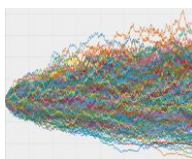


计算科学：模拟

利用电子计算机对复杂现象进行仿真取代实验；

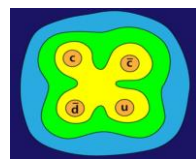


核试验 蒙卡模拟

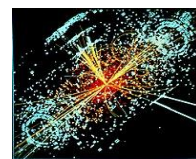


数据科学：数据

数据密集的科研范式，基于大数据分析并得出结论；

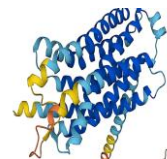


四夸克态 Higgs



智能科学：智能

依靠AI训练对复杂不确定性系统优化，从数据中发现规律；



蛋白质 **新粒子？新规律？**



第一范式

第二范式

第三范式

第四范式(当前)

第五范式

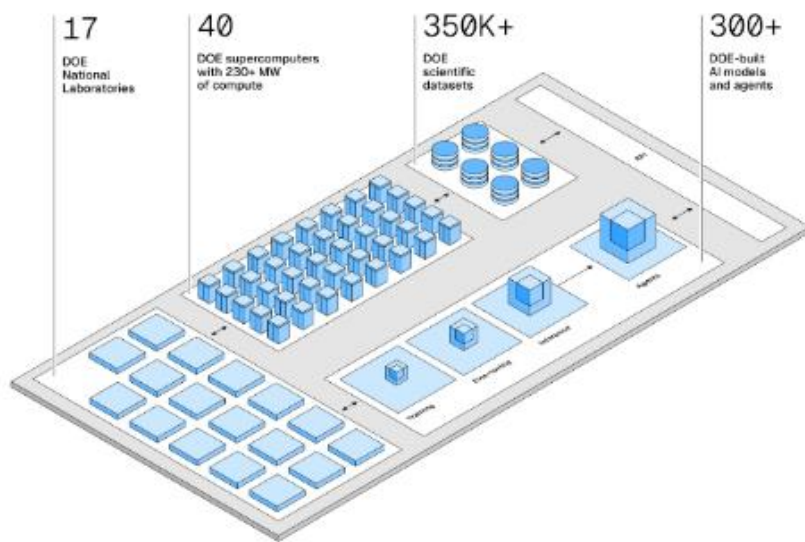
- 不同范式之间不是替代关系，而是**互为补充，协同发挥作用**。

为什么需要人工智能平台？



人工智能平台是推动AI4S规模化应用的关键基础设施。

美国创世纪计划：整合国家级科研资源



DOE 官方资源示意图；图中规模不等于已完成接入量

The Genesis Mission will build an integrated AI platform to harness Federal scientific datasets
——美国“创世纪计划”

01 资源互联

数据、算力与设施分散
统一资源发现与访问接口
降低跨设施协同门槛

03 流程贯通

实验与计算脱节，反馈滞后
编排数据处理、推理与实验
支持分析结果及时反馈实验

02 能力复用

重复开发，领域经验难沉淀
共享模型、软件与领域技能
让成果服务更多科研任务

04 可信治理

版本不清，结果难以复查
记录数据、模型与执行过程
支撑结果溯源与科学验证

fragmentation of resources and research efforts, difficulties in accessing computational resources and datasets ——欧盟《科学人工智能战略》

“We needed a whole ecosystem of software, infrastructure, and community.”

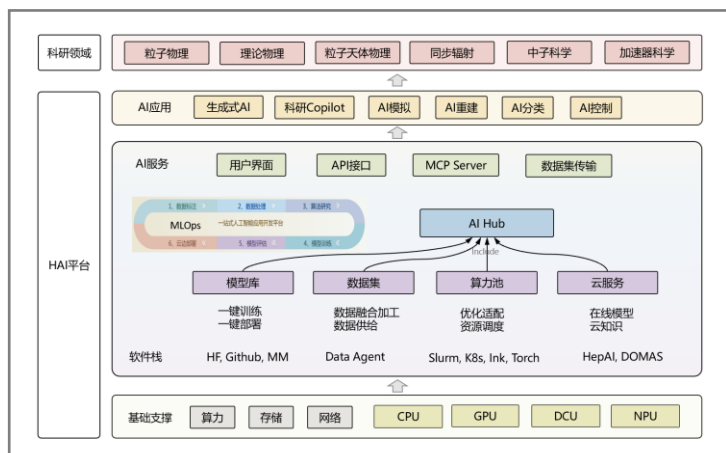
——伯克利国家实验室 Zhengde ZHANG

Part 2

高能AI平台



面向粒子物理、粒子天体物理、同步辐射、中子科学、加速器等多领域，连接模型、数据、算力、专业知识、科研软件 and 智能体，推动AI成为可复用、可共享的**新型科研基础设施**，加速科学发现。



🚀 平台价值

- **承载AI模型/工具**：已支持领域10+模型/工具。
- **建设专用AI算力**：40卡GPU+48卡DCU+本地模型算力资源
- **打通数据通道**：文本、语料、科学等10+AI数据集
- **赋能智能应用**：6+AI应用。

The HepAI Platform

<https://ai.ihep.ac.cn>



HepAI

模型 数据集 算力 应用 智能体 文档手册 EN 设置 用户

Next-gen AI Infrastructure

HepAI

高能人工智能平台

连接模型、数据、算力、专业知识、科研软件和智能体，推动AI成为可复用、可共享的新型科研基础设施，加速大科学装置科学发现。

DrSai物理分析智能体 →

Rongzai中子数据智能体 →

HaiAcademic学术助手

AI算力集群

实时数据处理

智能模型训练

高性能计算



34 ↑

可用模型



5 PFlops

算力



17 ↑

智能体/应用



9,579 ↑

用户



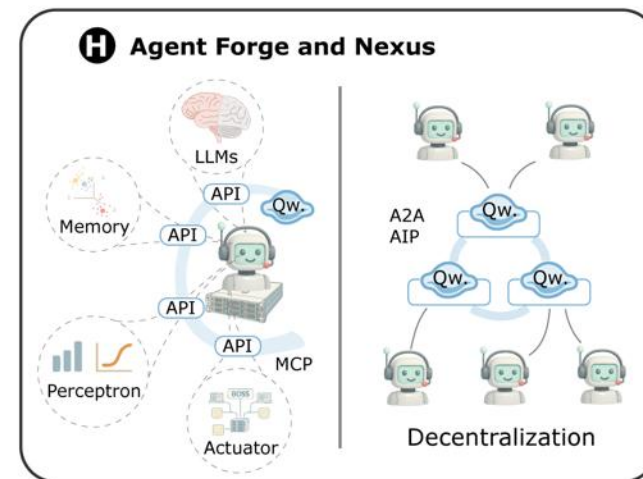
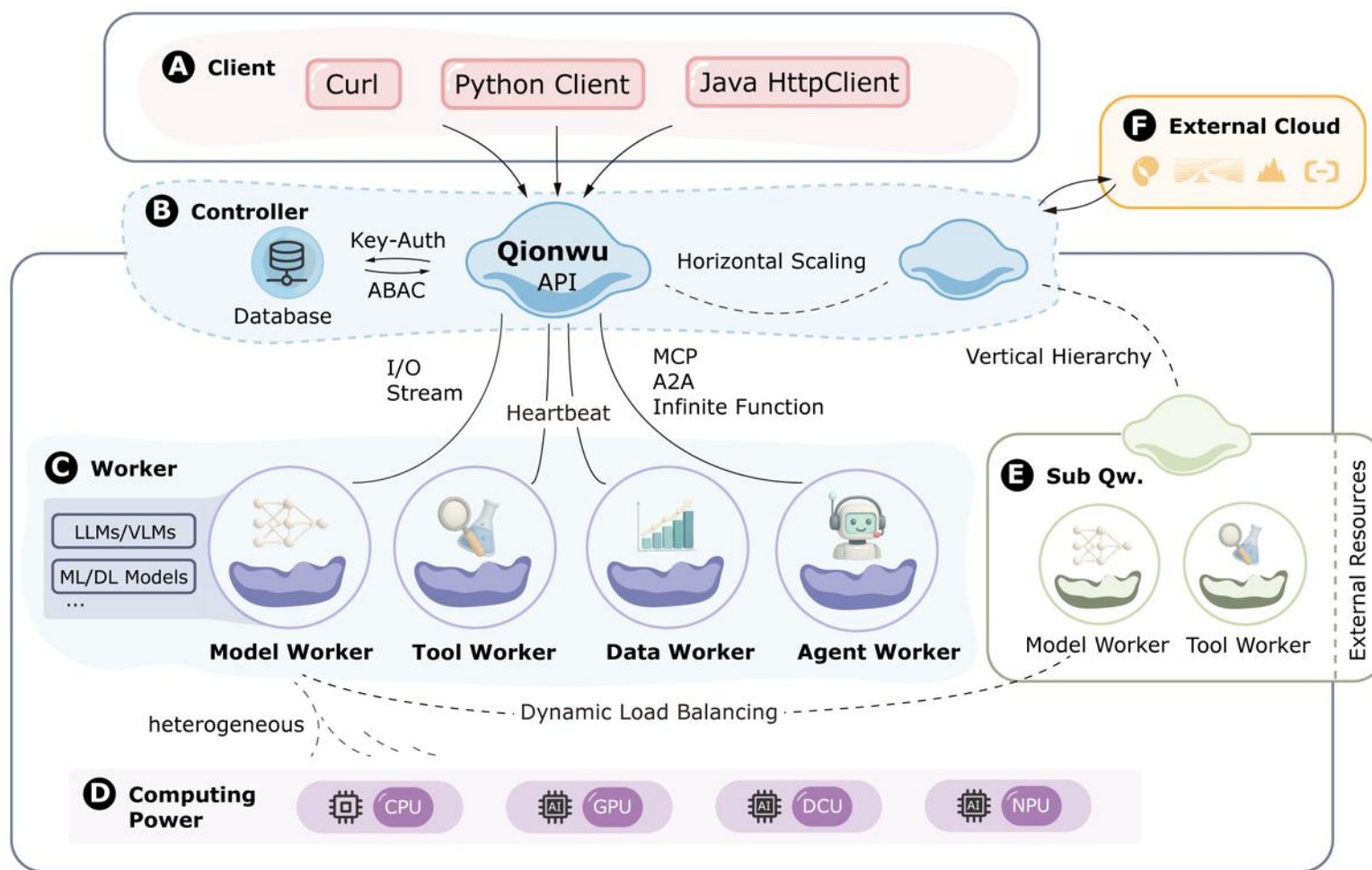
31,211 次

累计浏览量
自2025/11/19起

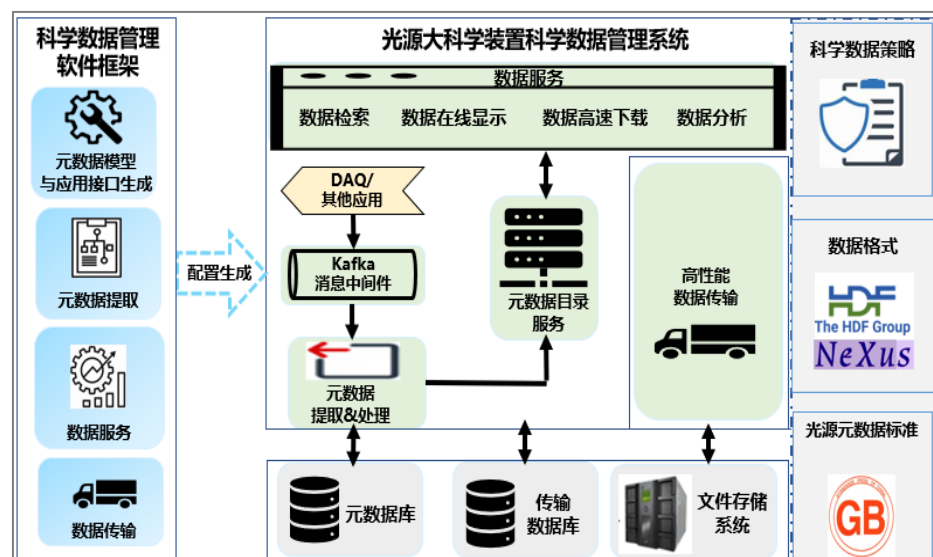
模型基础设施：DDF



分布式部署框架 (HaiDDF) 将人工智能要素 (如：**模型、科学工具、数据、智能体**等) 转换为可发现、可接入、可交互和可重用的资源，打破信息孤岛。



A **large-scale facility data system** refers to the comprehensive framework and mechanisms for the collection, storage, processing, analysis, and utilization of data within large scientific facilities. It encompasses the entire lifecycle of data, from its generation to its final application, and ensures data quality, integrity, security, and availability.



Domas Architecture



Fusion Data for AI Scientist

- Agents need to find data, understand metadata, trace provenance, access right files, and produce reusable results.

Providing AI elastic computing services on-demand to eliminate resource bottlenecks

Hardware Power:



Deployment Flexibility:

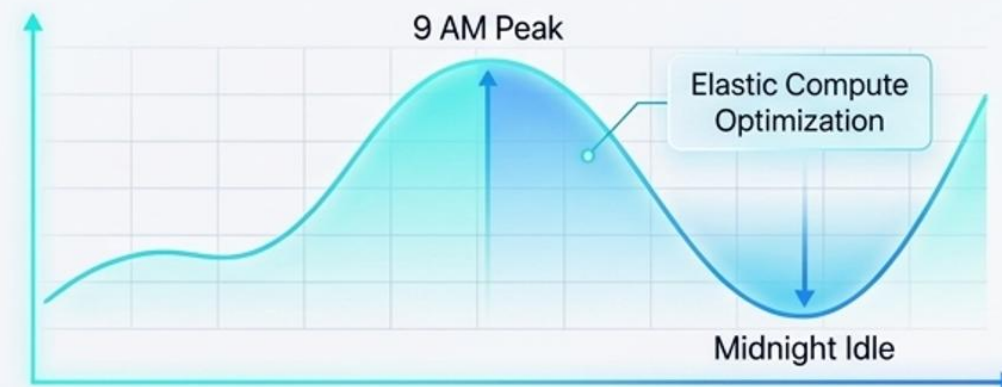
GPU: Single-machine, single-card training and deployment.

DCU: Large-scale model training.

NPU: Efficient model deployment.

Supports diverse workflows: EShell, Jupyter, Online VSCode, and VNC.

Compute Utilization Over Time



智能体基础设施：OpenDrSai

<https://opendrsai.ihep.ac.cn/>



- AI 智能体：智能体广场、精炼技能、MCP 等
 - 为智能体、技能、MCP 及其他 AI 生态服务提供接入能力
 - 随处运行：多平台、桌面端、文本用户界面（TUI）、网页与应用（APP）
 - 整合模型、工具、数据与算力资源

Agent Hub

OpenDrSai > Agent Square

TOTAL AGENTS
4

OFFICIAL
3

REMOTE
1

CUSTOM
0

All Local Remote

Remote
DocMaster
juzy@ihep.ac.cn
V1.0: A professional Word document processing tool, supporting the uploading, analyzing, editing...
[Start Chat](#)

Local
DrSai Neutron
xiongd@ihep.ac.cn
A general assistant for Neutron.
[Start Chat](#)

Local
DrSai Synchrotron
xiongd@ihep.ac.cn
A general assistant for Synchrotron.
[Start Chat](#)

Local
Rongzai Agent
wanghao96@ihep.ac.cn
Rongzai-Your neutron scattering data analysis assistant by CSNS data analysis group
[Start Chat](#)

Skill Hub

TOTAL SKILLS
43

TOTAL DOWNLOADS
555

TOTAL COLLECTIONS
2

All 计算中心服务 自动化 代码开发 科研 办公 LHAASO JUNO HEPs CSNS ALICPT

travelassistant-1.0.0
v1.0.0
AI智能旅游攻略助手。当用户需要规划旅行、查询目的地信息、获取出行建议时...
0 1

gitmind-1.0.2
v1.0.0
Use this skill whenever the user mentions GitMind, Gitmind, or localized GitMind...
0 0

ceshi
测试 skill 加载流程。当用户提到「测试 skill」[ceshi skill] 时使用。
1 0

revealjs-skill-main
v1.0.0
Create polished, professional revealjs presentations. Use when the user asks t...
0 0

Multi-platform agent harness

YOUR AGENT, EVERYWHERE

Your agent, ready to work for you anytime

Chat in the browser or choose the client that fits your workflow. Your sessions stay connected.

[Start chatting](#) (Use in your browser)

[Get the apps](#)

WebUI Use in your browser [Open](#)

Windows Connect local workspaces [View](#)

Android Work and follow progress [View](#)

Terminal UI Terminal vbe coding [View](#)

macOS Apple silicon [View](#)

Agent Client for Windows/Macos

File Edit Layout Help

OpenDrSai Desktop

hi ready

YOU hi

OpenDrSai - drsai 已完成 · 8 秒 过程

Hi there! I'm OpenDrSai, your intelligent assistant for software engineering and data analysis tasks. I can help you with things like:

- Coding & debugging** — exploring code, fixing bugs, refactoring
- Data analysis** — crunching numbers, building visualizations
- Document generation** — Word docs, spreadsheets, presentations
- Research tasks** — searching codebases, verifying information

It looks like you're currently in the `D:\work\projects\drsai` project directory. Based on my memory, I've worked with you on **Windows desktop release publishing via Alibaba Cloud OSS/CDN** before.

What would you like to work on today?

YOU 我有哪些技能

OpenDrSai - drsai 已完成 · 12 秒 操作 过程

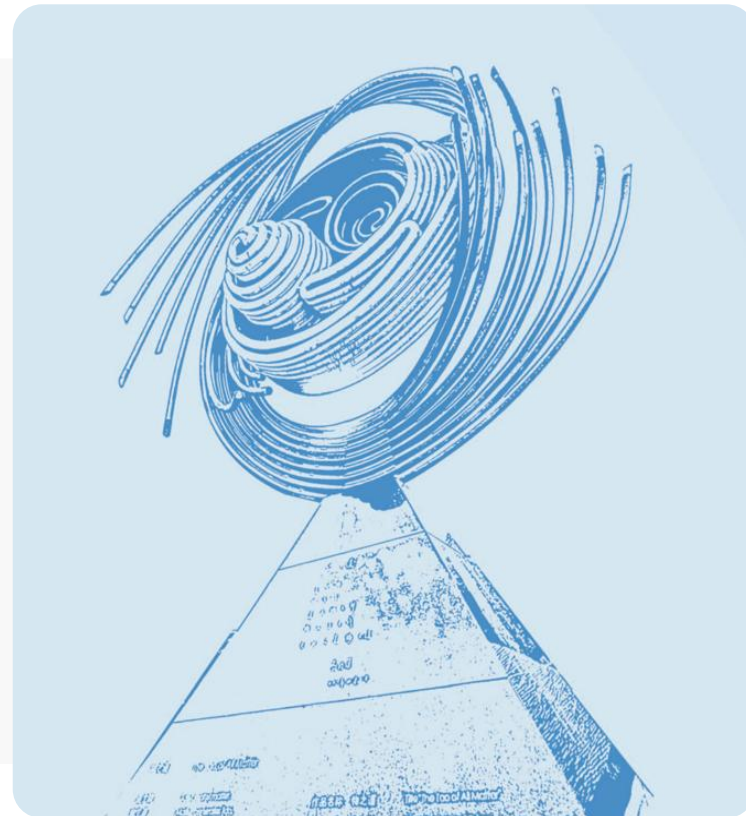
Let me check what skills are available for you.这个技能是关于系统设置的，不适合回答你的问题。让我直接根据系统已

[Ask OpenDrSai...](#)

KB OpenDrSai - gpt 5.6 luna - High - Normal

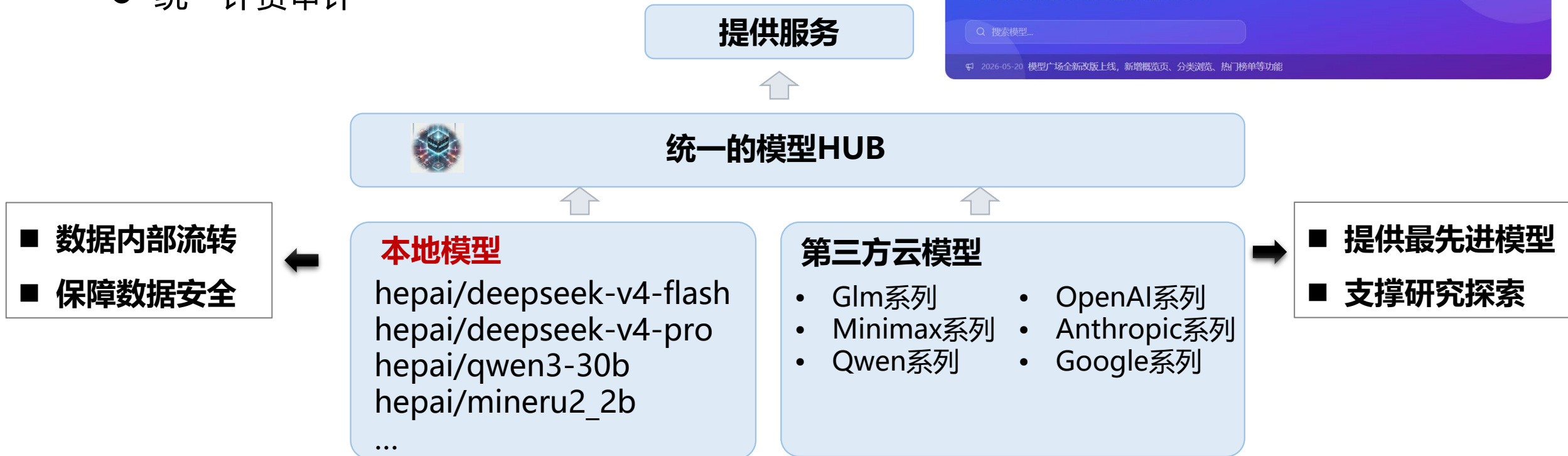
Part 3

平台应用服务



- 基于模型基础设施的**大模型词元 (Token) 服务**，提供统一的API接口和计费
- 方案：私有化本地部署+第三方云模型协同
 - 模型**开箱即用**
 - 并发高
 - 统一计费审计

<https://ai.ihep.ac.cn>



HepAI Model Plaza

IHEP LLM Startup

May 22, 2026

¥1000 LLM credit has been issued to your platform balance

65+ Available Models

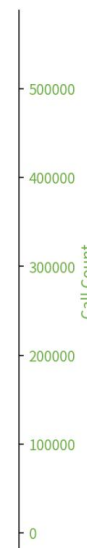
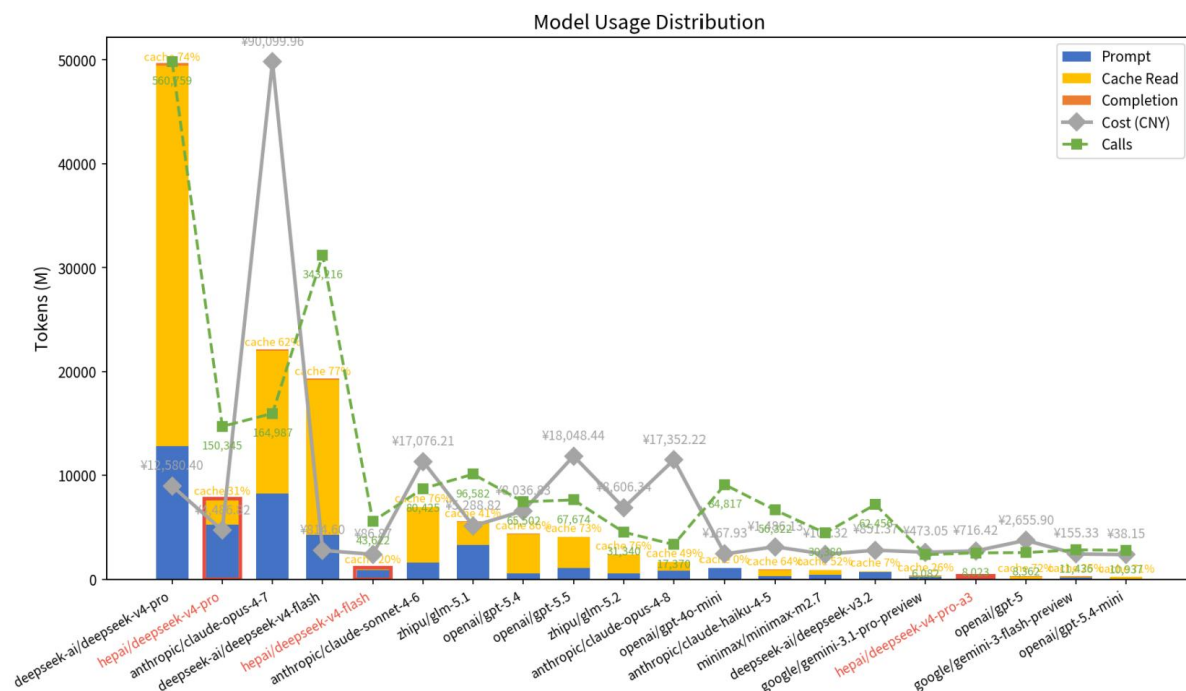
10+ Providers

[Browse All](#)

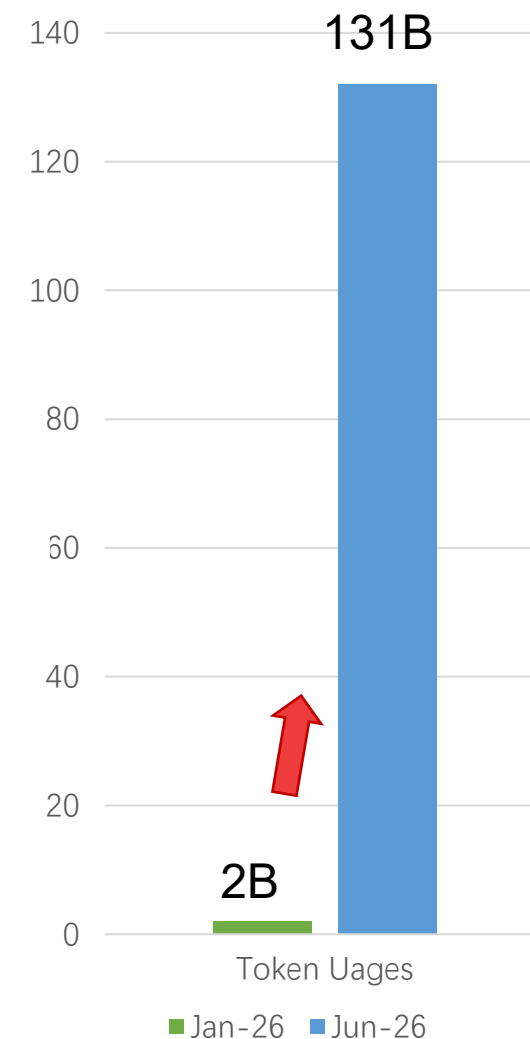
[Docs](#)

- For every employee, post-doc and students.
- Equivalent to ~55M tokens of deepseek-v4-pro

Static data (2026-06-01 00:00 UTC -- 2026-06-30 23:59 UTC)



- Users: 1335
- Requests: 2.3M
- Tokens: 131B



Already equivalent to a large-scale scientific computing service

核心指标：

- (1) **TTFT, 首Token延迟**, 即流式输出下用户发起请求到返回一个Token的时间。
- (2) **TPS, 每秒输出Token数**。

上述指标在不同的并发数（服务端同时处理的请求数量）下表现差异大。

测试方案：

- 1.输入1k输出1k（Chat标准）；
- 2.要求输出撞max_tokens，消除噪声；
- 3.关闭thinking防止污染TTFT（首字延迟）
- 4.提示词注入UUID nonce，破kv-cache防止TTFT失真
- 5.设置warmup，消除冷启动+连接池开销

AI模型服务：模型并发测试结果



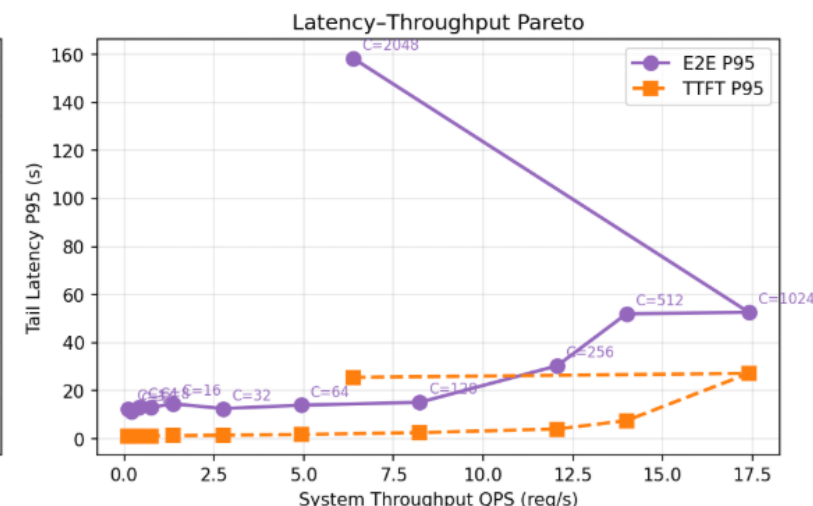
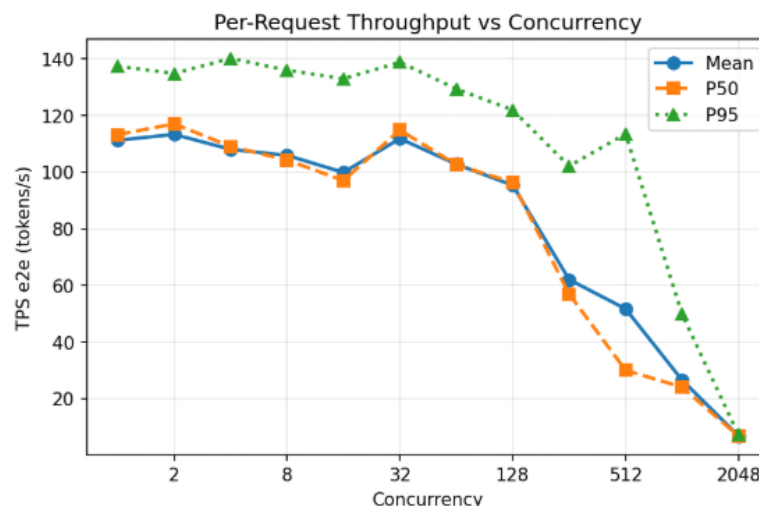
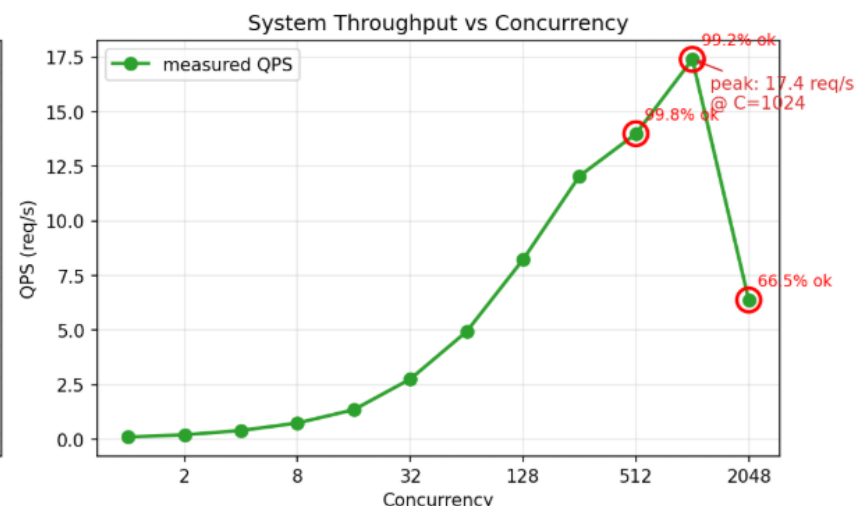
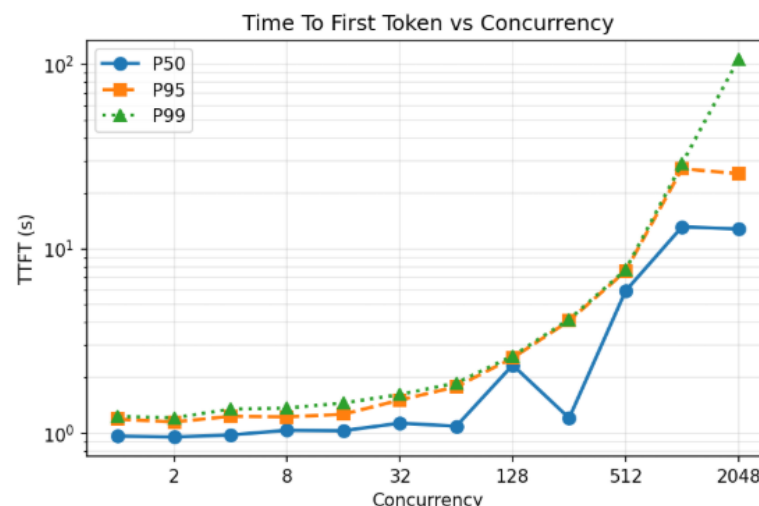
配置：

- DDF生产环境（HAI平台）
- 云服务
- Deepseek-v4-flash模型

结果：

- 1-512并发高性能可用：TTFT在秒量级，TPS约120
- 1024并发时TTFT上升明显，可用。
- 2028并发时，请求成功率79%，部分请求无响应。

Concurrency Sweep — ddf_sweep_20260518_021810





桌面端

直接开始对话 →
(在浏览器中使用)

下载客户端



终端UI

网页UI

```
加载个人及系统配置文件用了 1149 毫秒。
(base) PS C:\Users\26364> opendrsai
<frozen runpy>:128: RuntimeWarning: 'drsai.backend.run_cli' found in sys.modules after import of pa
tion of 'drsai.backend.run_cli'; this may result in unpredictable behaviour
🔥 OpenDrSai
Website: https://opendrsai.ihep.ac.cn/ · Guide: https://note.ihep.ac.cn/s/QgtE3Nlx2
Version: 1.5.6
📄 Memory ~\drsai\workspace\runs\xionddb\configs\MEMORY.md (15/15 entries)

• connected • xionddb @ deepseek-v4.1-flash • 13 tools • 📁 C:\Users\26364 • any-path • any-cmd
agent.ready: agent ready

> | Send a message • @ files • / commands • Tab complete • Ctrl+0 newline
```

使用开源DeepSeek Harness



使用步骤

1 安装并打开
下载并安装 DeepSeek Harness...

2 进入模型设置
点击左下角“设置”，选择“模型”，...

3 填写 HAI 配置
依次填写 API 密钥、API 地址和 o...

4 添加模型
加入 deepseek-v4-flash 和 deep...

API 密钥
sk-xx
从 HAI 平台获取 [🔗](#)

API 地址
https://aiapi.ihep.ac.cn/apiv2

API 协议
openai-responses

点击复制图标，将左侧三项依次填入右图对应位置。

设置

通用设置

模型

插件

Agent 预设

hepai 自定义

API 密钥

已配置——输入新值可替换

自定义设置

显示名称

hepai

API 地址

https://aiapi.ihep.ac.cn/apiv2

API 协议

openai-responses

模型目录

已自定义模型目录

恢复默认模型

获取可用模型

deepseek-v4-flash

显示名称

deepseek-v4-pro

显示名称

添加模型

取消

保存

4 选择模型并开始对话
在消息发送框切换刚添加的模型

deepseek-v4-flash

deepseek-v4-pro

数据不出所

需要指定高能所内部模型时，为模型名称添加 hepai/ 前缀。

hepai/deepseek-v4-flash

在容器中使用智能体



通过智能体使用大模型

基于AI算力服务与AI模型服务，AI平台可支撑并提供先进的AI智能体服务

- Openclaw、Hermes、Claude Code、Opendrsai等

算力 应用

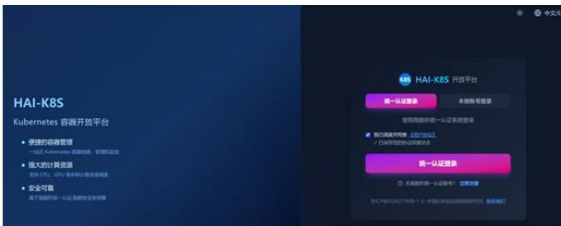
HepAI智算-K8S入口
k8s-ai.ihep.ac.cn

<https://k8s-ai.ihep.ac.cn/apps>

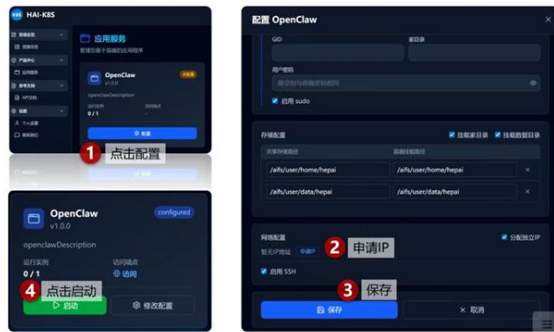
1. OpenClaw

持续运行在高性能集群上，执行 shell 命令、管理文件，并协调多步骤工作流程

第一步：登录HAI-K8S开放平台：<https://k8s-ai.ihep.ac.cn>



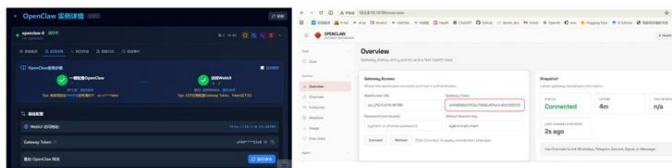
第二步：配置应用和启动容器



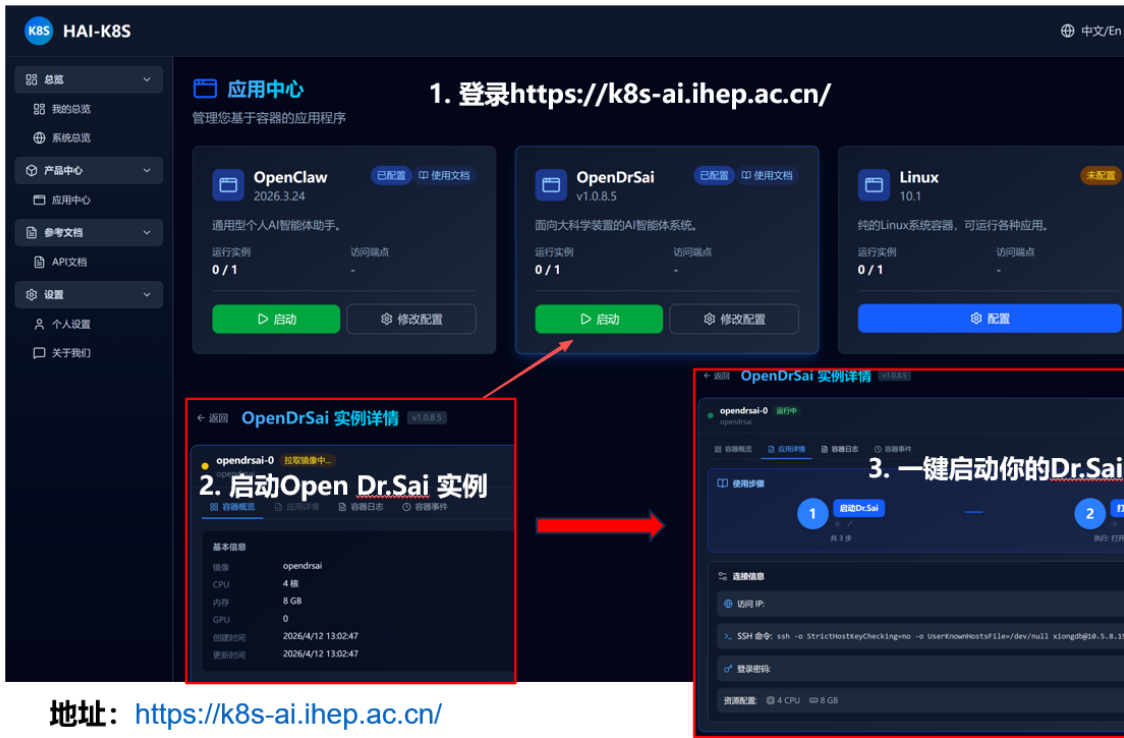
第三步：初始化启动OpenClaw



第四步：打开界面粘贴Gateway Token，并开始对话！



2. Open Dr. Sai通用智能体



地址：<https://k8s-ai.ihep.ac.cn/>

OpenDrSai容器内附带Claude Code

- 输入Claude启动
- HAI平台的BASE-URL和API-KEY已自动配置

```
zdzhang@hai-opendrsai-0:~$ claude
Claude Code v2.1.145

Welcome back!



Opus 4.7 (1M context) · API Usage Billing
/aifs/user/home/zdzhang

Tips for getting started
Run /init to create a CLAUDE.md file with instructions for Claude
Note: You have launched claude in your home directory. For the best experience, launch it in a proje...

What's new
Added `claude agents --json` to list live Claude sessions as JSON for scripting (tmux-resurrect, sta...
Added `agent_id` and `parent_agent_id` attributes to `claude_code.tool` OTEL spans, and fixed trace ...
Status line JSON input now includes GitHub repo and PR information when detected
/release-notes for more

> hello

• Hello! How can I help you today?

* Cogitated for 3s

> |
```

基于HAI平台的原版OpenClaw

k8s-ai.ihep.ac.cn



所内专属独立IP



基于容器的环境隔离

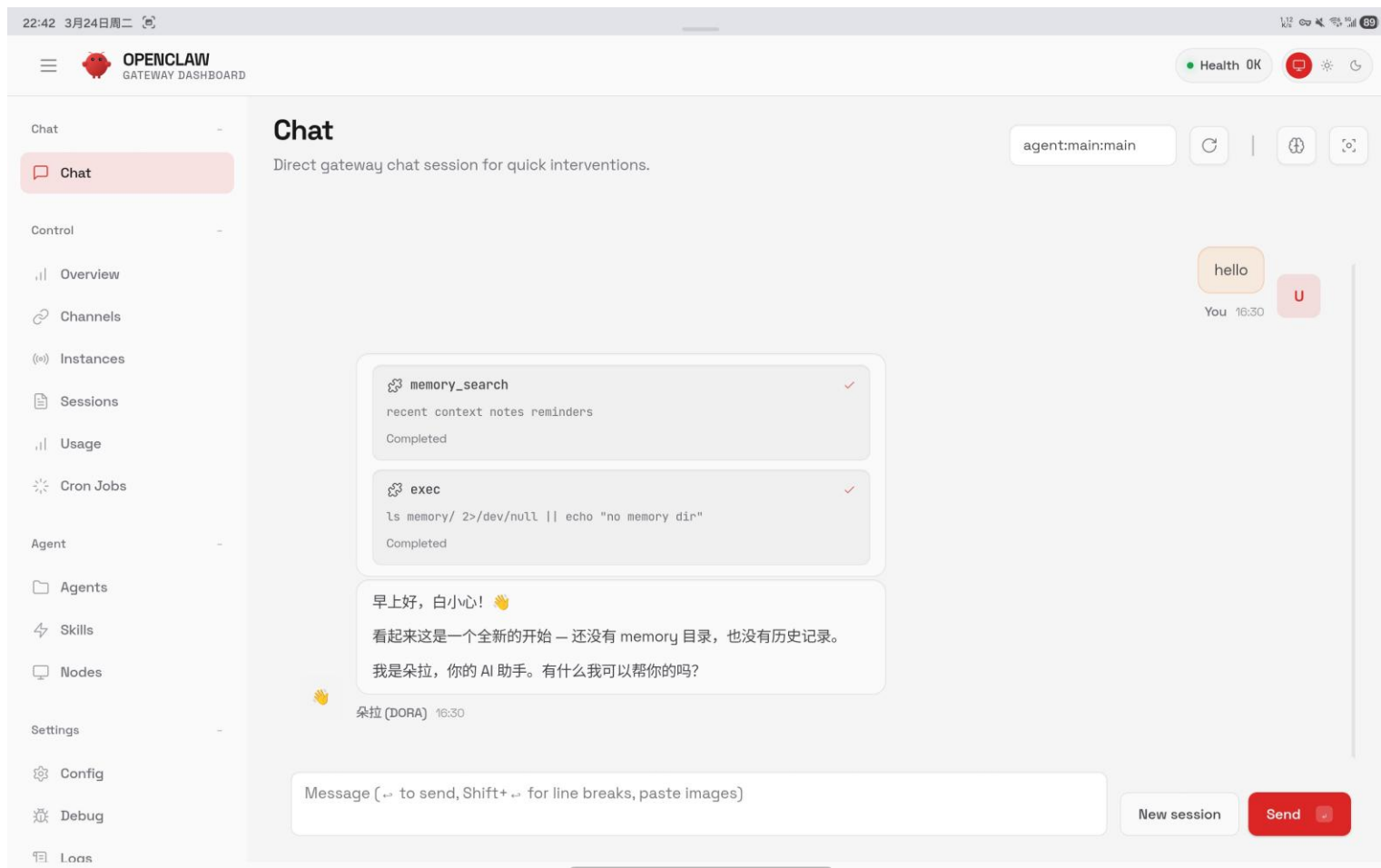


数据持久化

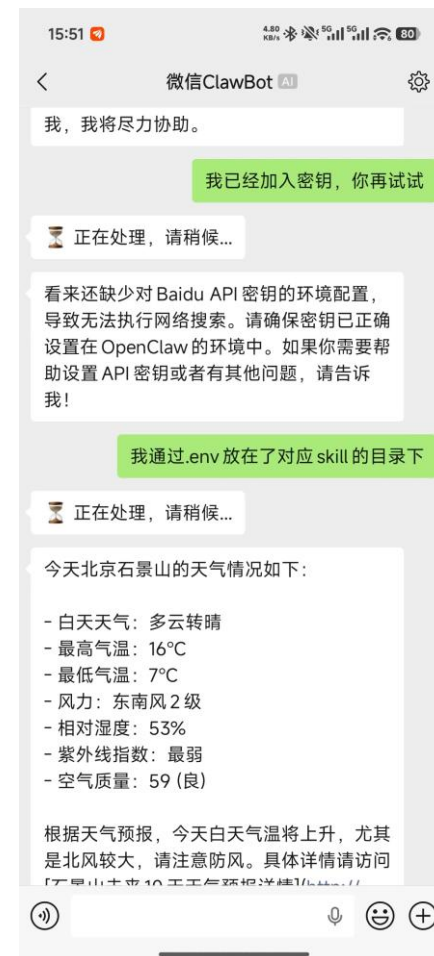


本地模型和云端模型无缝切换

http://10.5.8.xx:18789/



WebUI Controller

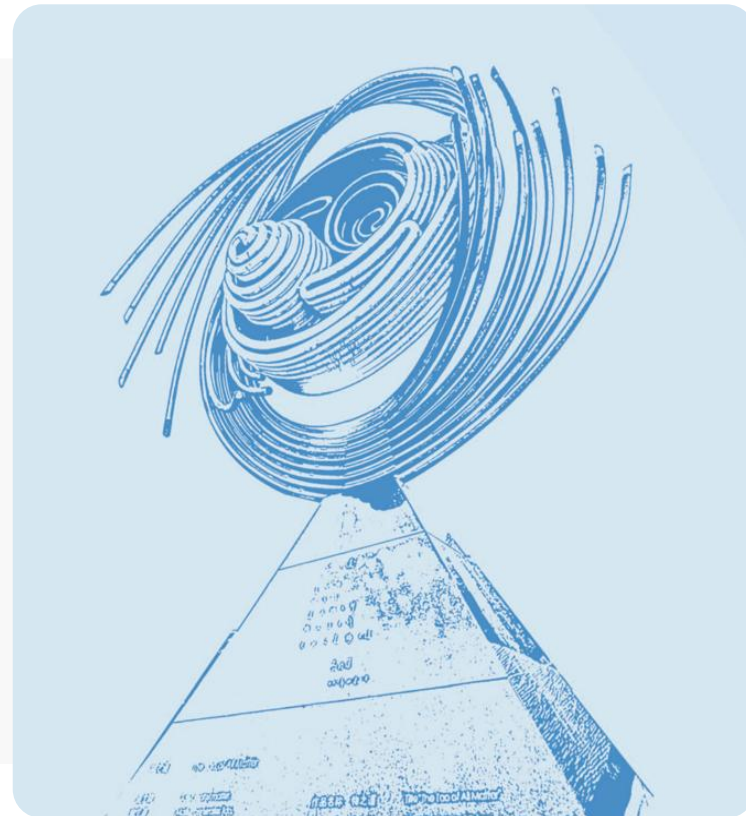


接入微信等app

Zhengde ZHANG

Part 4

总结



- 划定底层红线。
- 坚守学术道德。

- 警惕“认知卸载”
- 注重“批判思维”

- 知晓技术不成熟性。
- 预防重要文件“误操作”。
- 警惕功能插件 (skills) 投毒。

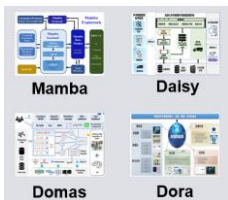
到2030年，期待人工智能能够赋能研究所产出 **N 项** 重大成果

技术支持

大装置数据体系



◆ 发展数据科学，**数据软件自主可控**



◆ 建成**PB级**AI就绪的领域科学数据库/集

◆ 推动科学数据共享利用规范体系（光源/中子源）

模型系统



- ◆ **开源模型**的本地化应用和优化
- ◆ 推进特定任务的**专用AI模型、大小模型协同方法**
- ◆ 探索理解科学数据、掌握物理规律的**科学(大)模型**



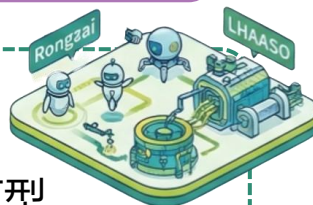
提供数据

提供模型

运行支撑

运行支撑

智能体与工具链



- ◆ **物理分析智能体**从“原型系统”走向“**实用**”
- ◆ **数据智能体**反哺数据体系
- ◆ 推动一批智能体落地应用（JUNO、HEPS、CSNS、LHAASO等）
- ◆ 持续续发展领域工具和Skills

运行支撑

AI基础设施

- ◆ 持续提升**算力能力**，确保**数据安全**
- ◆ 优化**模型基础设施**，本地/云相结合
- ◆ 上线**AI-Ready数据** 服务
- ◆ 与外部设施联动和互联互通

Related Links



HepAI Platform: <https://ai.ihep.ac.cn>

Open Source SDK: <https://github.com/hepai-lab/hepai>

Agent Harness: <https://opendrsai.ihep.ac.cn>

HEP·Xiwu LLM: <https://github.com/zhangzhengde0225/Xiwu>



HepAI Wechat Group



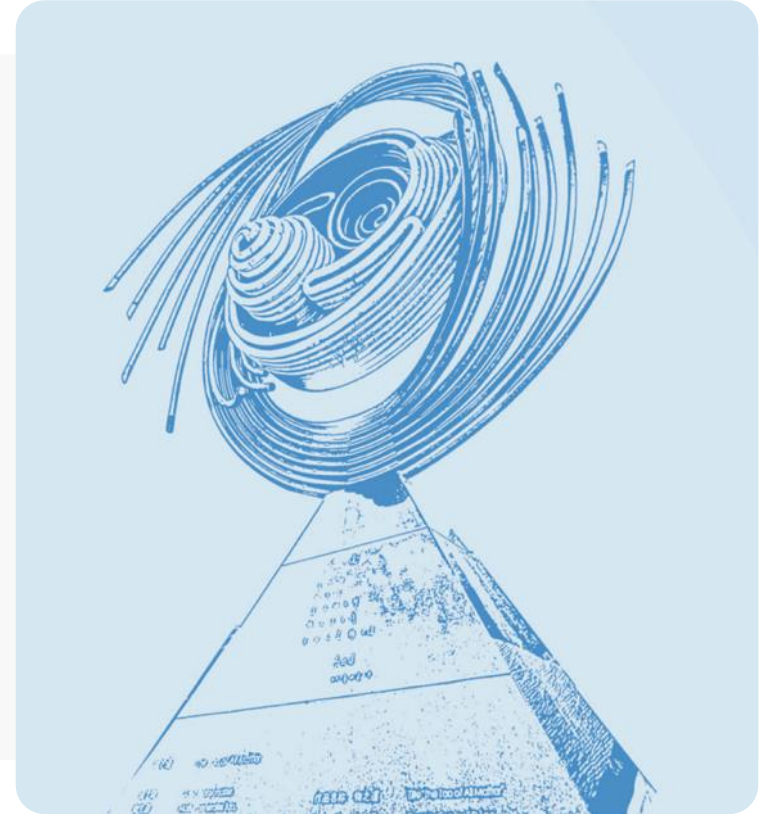
- **Let's discuss AI in high-energy physics — we welcome your input!**

- Zhengde Zhang (zdzhang@ihep.ac.cn)
- Institute of High Energy Physics
- National HEP Data Center

Thank you



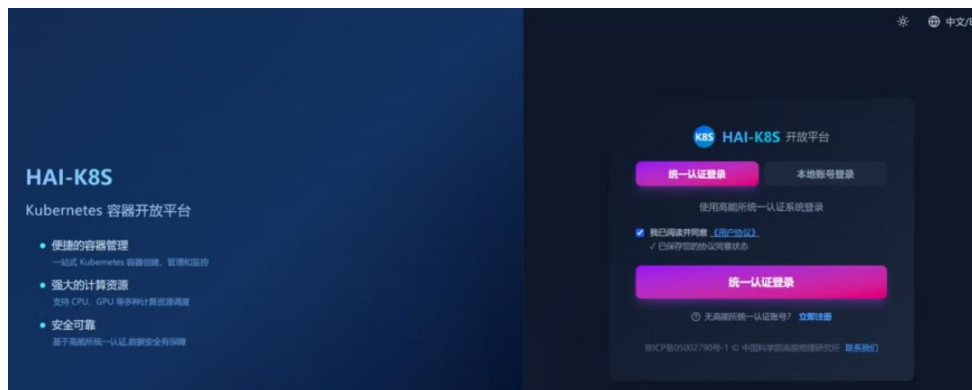
Backup



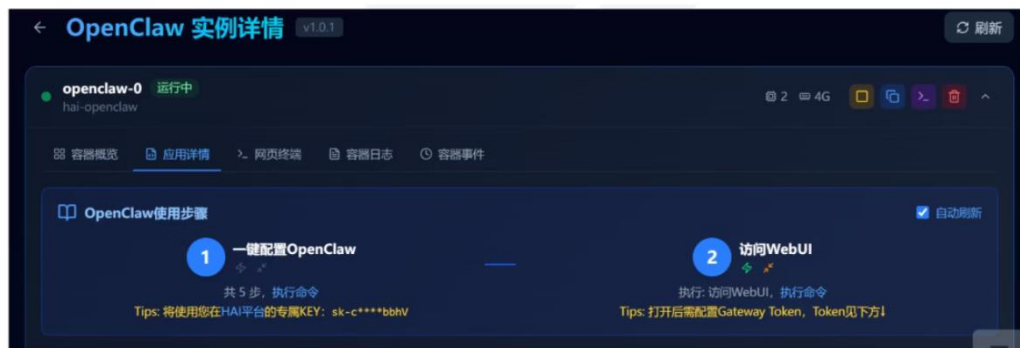
OpenClaw使用方法



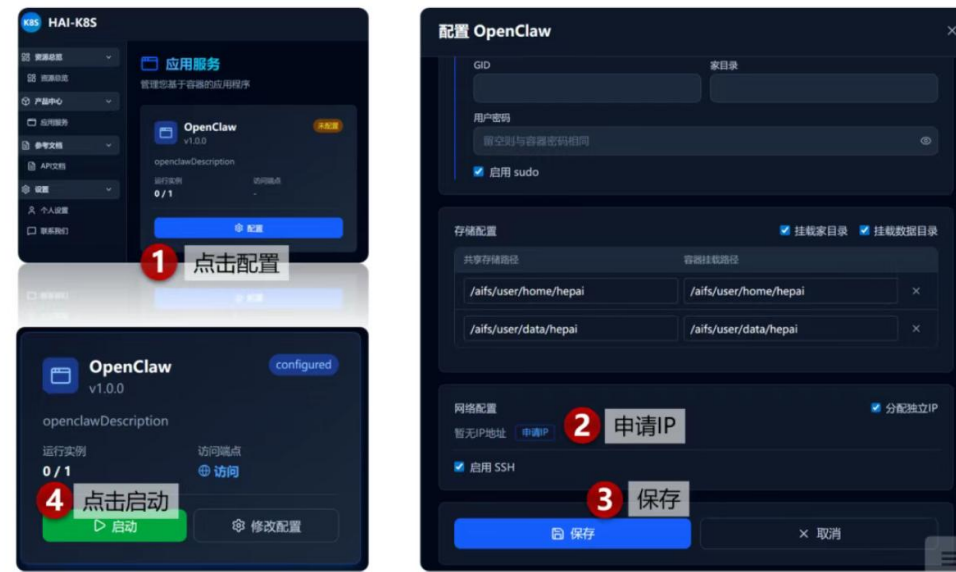
第一步：登录HAI-K8S开放平台：<https://k8s-ai.ihep.ac.cn>



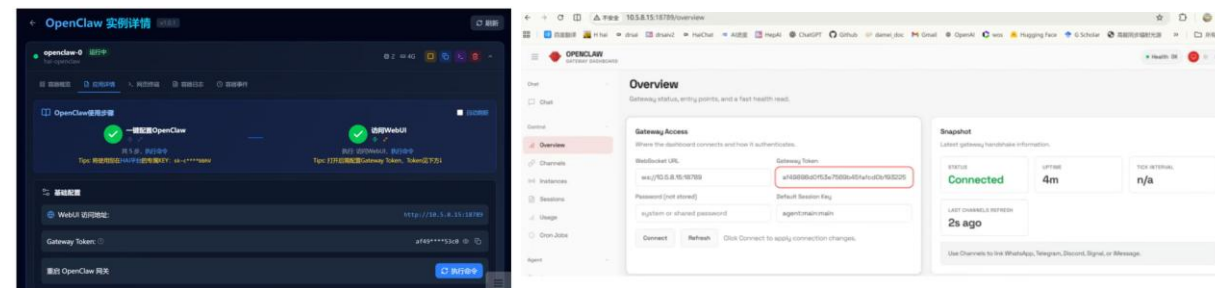
第三步：初始化启动OpenClaw



第二步：配置应用和启动容器



第四步：打开界面粘贴Gateway Token，并开始对话！



【详细教程】HAI平台一键拥有原版专属OpenClaw：<https://note.ihep.ac.cn/s/h28FTEGfP>

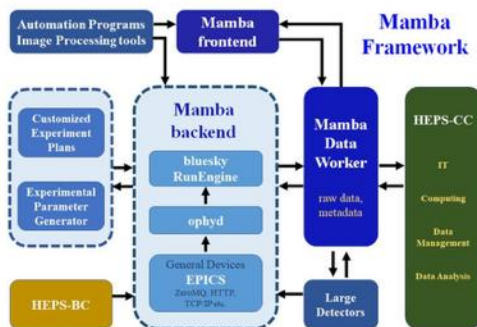
大装置数据体系进展--数据软件



- 当前状态：覆盖数据生产到使用的全生命周期的框架/工具/软件的**初版得到应用**

01 数据采集与解析

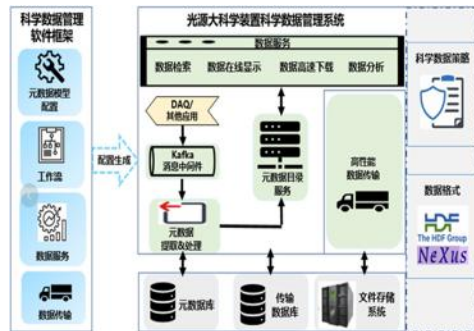
实验控制和数据 采集框架Mamba



- ❑ 复杂实验过程控制
- ❑ EB级大数据采集
- ❑ 高效在线解析

02 数据存储与管理

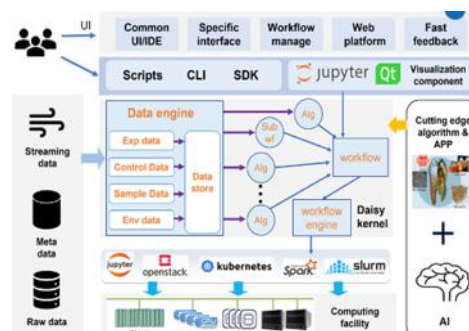
数据管理与服务 框架Domas



- ❑ EB级大数据存储
- ❑ 复杂的元数据管理
- ❑ 标准化数据格式

03 数据处理与分析

科学数据处理软 件框架Daisy



- ❑ 多方法学集成
- ❑ 高通量数据I/O
- ❑ 学科软件及界面

04 数据清洗与标注

数据融合智能体 Dora



- ❑ 高精度自动标注
- ❑ 互补性数据融合
- ❑ 信息理解和辅助

- 下一步考虑：优化和推广数据软件在各装置中**深入应用**。

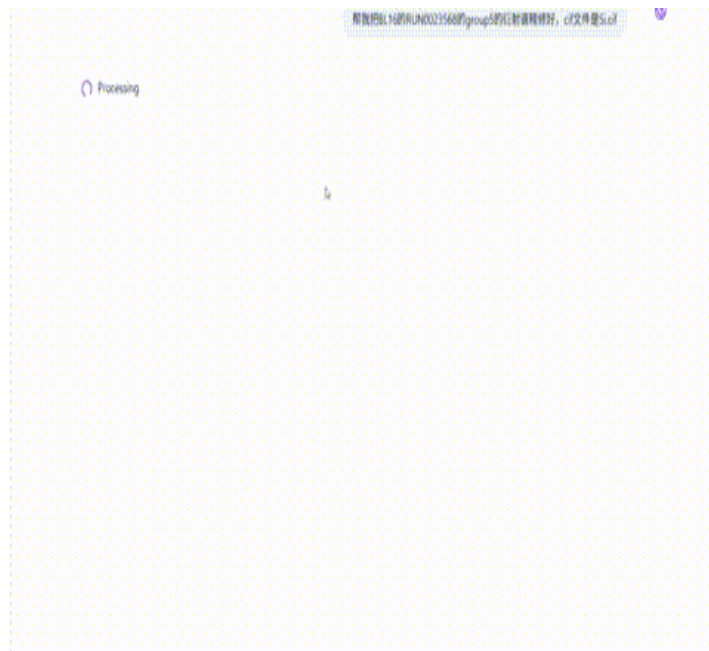
平台层成果：衍射数据联合精修智能体



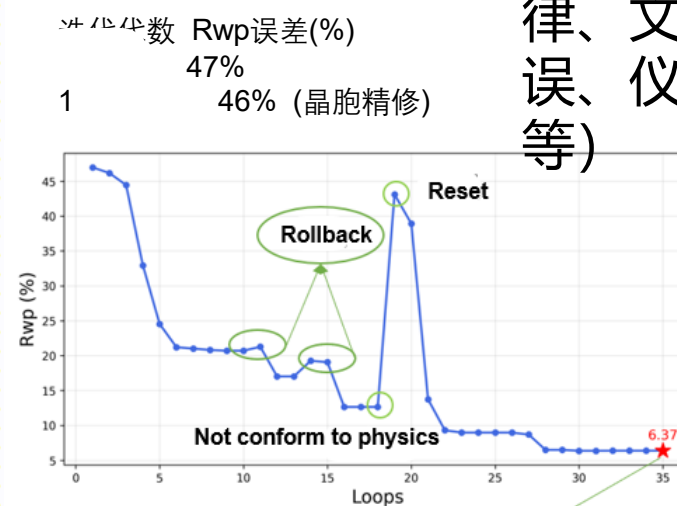
ArXiv: <https://arxiv.org/abs/2605.13911>

- 面向X射线衍射和中子粉末衍射联合精修的AI智能体
- 从实验数据和初始结构出发，自动规划任务、使用工具、验证结果、优化方案，实现结构精修。

提示词: 使用task_003文件夹下的文件LaB6.cif, LaB6.txt, initial.instprm 进行XRD精修



- 自动优化精修策略
- 自动回退（自主发现不符合物理规律、文件参数错误、仪器参数错误等）



结果（之一）：6.37% (智能体) vs 7.78% (人类专家)

- 首次复杂自动精修达到人类专家水平，效率提升5倍，国际领先。



①课题负责人关联课题

https://ai.ihep.ac.cn

HepAI

个人中心
管理您的账户、API密钥、用量统计等

我的课题
管理您的课题信息

从Dataway导入 + 手动添加课题

重要提示
添加和导入课题后，课题信息将用于平台计费与权限控制，请确保信息准确。课题需经审核通过后方可使用，审核期间请耐心等待。

科研服务-人工智能服务 已通过 已禁用
课题号: [ID] 负责人: 张正德 机构: 中科院高能所
负责人邮箱: zdzhang@ihep.ac.cn

财政-基本科研费-从0到1项目-AI大模型驱动高能物理科学发现的可行性研究 已通过 已禁用
课题号: [ID] 负责人: 张正德 机构: 中国科学院高能所
负责人邮箱: zdzhang@ihep.ac.cn

财政-基本科研费-从0到1项目-AI大模型驱动高能物理科学发现的研究项目预算 已通过 已禁用
课题号: [ID] 负责人: 张正德 机构: 中科院高能所
负责人邮箱: zdzhang@ihep.ac.cn

③课题负责人审核

我的审核
查看您需要审核的事项

筛选条件
全部类型 全部状态 年/月/日 年/月/日 查询

审核记录
刷新

标题	类型	申请人	操作
使用课题充值申请 code: Hr-9****d9cb	激活码	zdzhang@ihep.ac.cn	同意 拒绝

② 用户申请充值额度

HepAI

个人中心
管理您的账户、API密钥、用量统计等

账户管理
管理您的账户余额、交易记录、调用记录

账户余额
¥ 3239.88
约 809,969,782 tokens
free
立即充值

课题充值
选择课题后,充值申请将发送给课题负责人审核确认。

充值金额(元)
100
50元 100元 200元 500元 1000元

申请人姓名
张正德

搜索课题
张正德

温馨提示: 请先前往「我的课题」中导入您的课题信息, 导入后即可在此搜索使用。

选择课题 (1个结果)

财政-基本科研费-从0到1项目-AI大模型驱动高能物理科学发现的可行性研究
编号: E3545PU2 | 负责人: 张正德 (zdzhang@ihep.ac.cn)
✓ 已选择
立即充值

④用户获取访问密钥

在任意客户端设置Base Url、密钥和模型即可访问

HepAI

API密钥
创建和管理您的 API 密钥

使用步骤
点击「创建新密钥」生成密钥
妥善保存密钥 (仅创建时可见)
Base URL: https://aiapi.ihep.ac.cn/api/v2

安全提示
请妥善保管您的API密钥。切勿在客户端代码、公开仓库或社交平台中分享。如怀疑密钥泄露请立即删除并重新生成。

创建新密钥

API密钥列表

序号	名称	密钥	过期时间	备注	操作
1	for_myself	sk-*****vXs3	2029-08-01		