



# Open Dr.Sai: 面向科学的自主 AI智能体

赋能高能物理，迈向大科学  
装置智能化新纪元

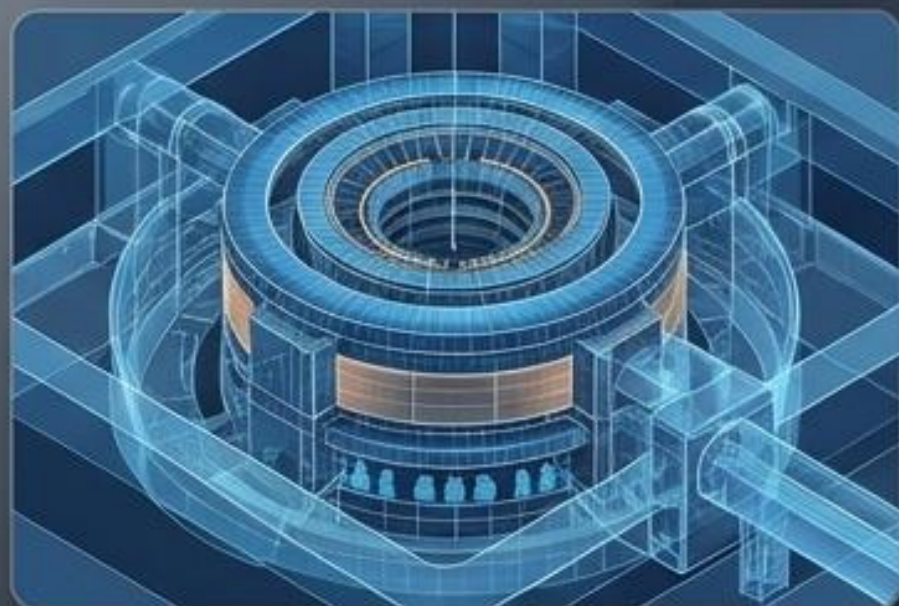
---

张正德， On behalf of Dr.SAI working group  
高能所计算中心/国家高能物理科学数据中心

注：本PPT为AI生成

# 维度诅咒：突破海量科学数据的认知瓶颈

**瓶颈：**大型科学装置产生了世界上最大规模的高质量数据集。人类研究人员面临严重的人力限制和“维度诅咒”，无法充分挖掘这些数据的全部潜能。



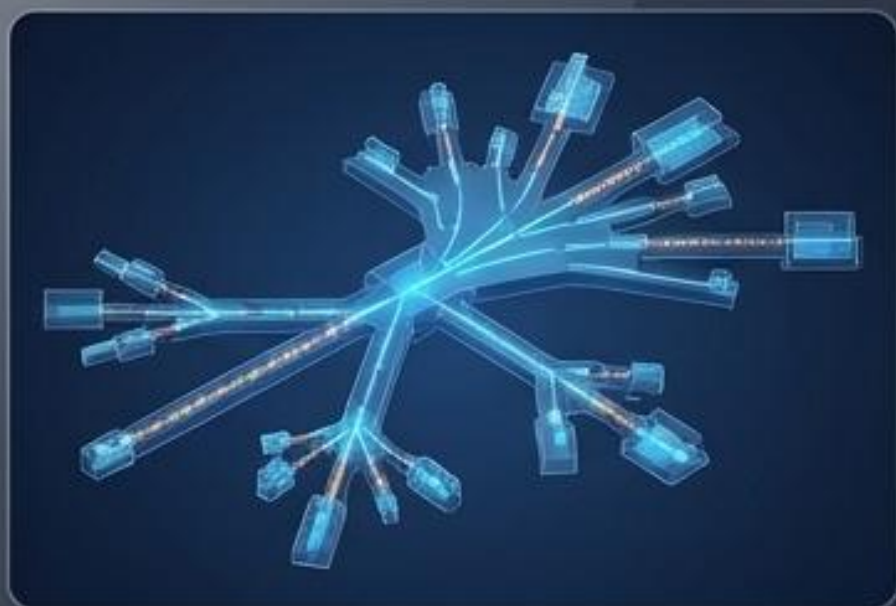
**BESIII** (Beijing Spectrometer III)

- 202米直线加速器
- 242米环形加速器
- 世界上最大的陶-粲数据样本



**HEPS** (High Energy Photon Source)

- 1.36km储存环
- 世界上最亮的第四代同步辐射光源之一



**CSNS** (China Spallation Neutron Source)

- 世界级大型中子散射多学科研究平台

**破局：**AI 提供了一条全新路径，绕过人类认知极限，从海量高维数据中提取隐藏的物理学规律。

# 科学AI的演进阶梯：从辅助工具到自主科学家

## Level 3: Autonomous Scientist (顶层 - Open Dr.Sai 目标)

- 提出原创科学假设并主动获取数据
- 自主调用特定工具并验证实验结果
- 创造全新的科学认知

## Level 2: AI as Reasoner (中层)

- 具备更高阶的智力容量
- 总结复杂模式并执行逻辑推理
- 生成“半成品”科学发现

## Level 1: AI as Assistant (底层)

- 克服人力和维度瓶颈
- 自动化重复性任务
- 执行高维数据降维

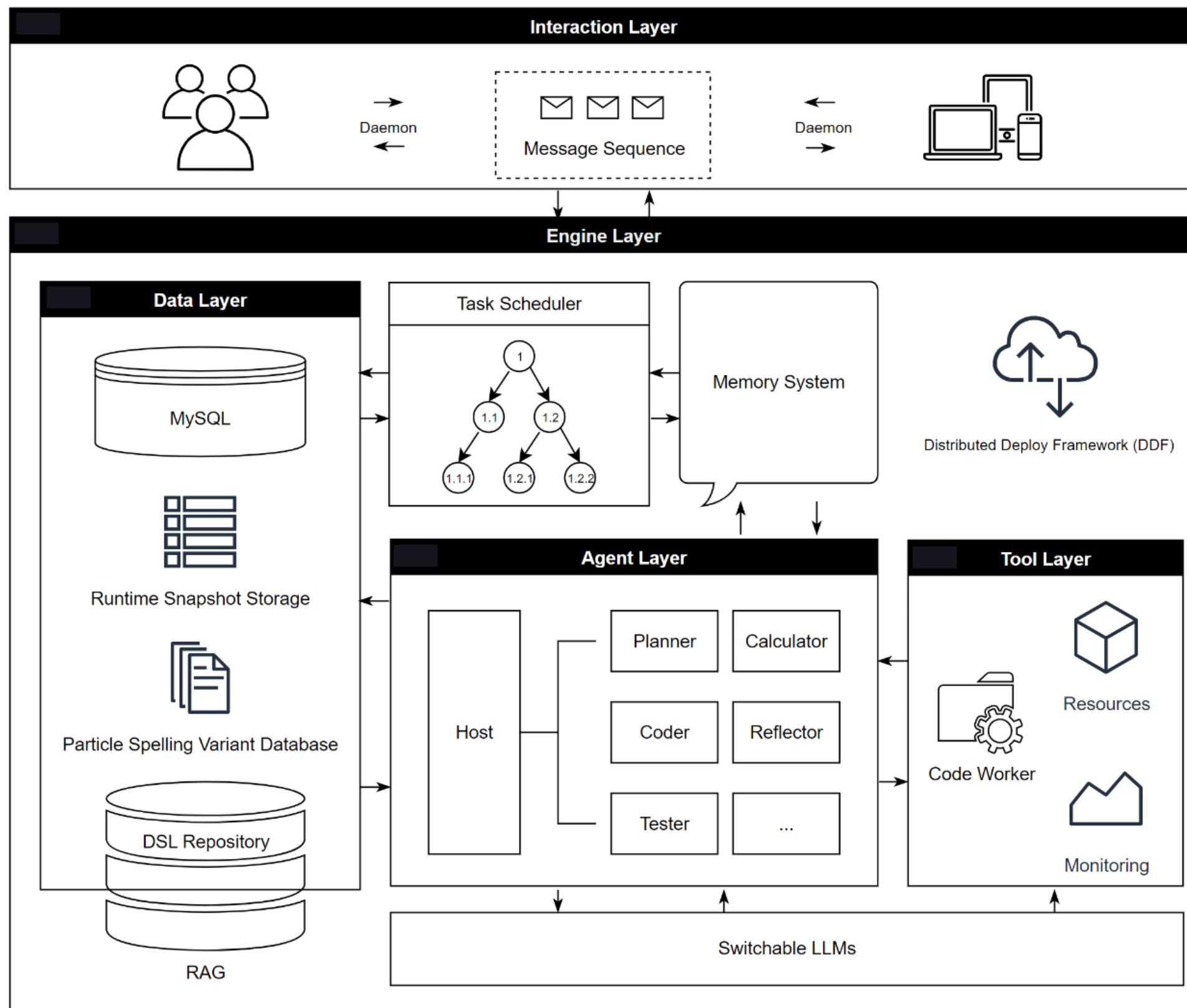


# Open Dr.Sai v2.0: 新一代多智能体系统概览

面向大科学装置的AI智能体系统。



# Dr.Sai 的架构



# 核心架构理念：驱动自主性的 BAMS 模型

去中心化、开源的多智能体枢纽，具备强大的权限验证与严格的人机交互护栏。支持 MCP, A2A, AIP 等无限函数协议 (IF)。



## B - Brain (大脑)

Large Language Models (LLMs) - 系统的认知与决策核心。

## M - Memory (记忆)

Persistent state retention - 持久化状态保留与上下文管理。



## A - Actuators (执行器)

Action outputs & tool invocation - 物理动作输出与领域工具调用。



## S - Sensors (传感器)

Perceptors - 智能感知输入，解析复杂环境与数据。

# 系统核心能力矩阵

## Continuous Evolution (持续进化)

- 数据飞轮系统 (Data flywheel)
- 运行时快照与动态标注



## Human-Computer Interaction (人机交互)

- 高可用基础接口
- 长周期任务的上下文交互交互机制



## Complex Tasks (复杂任务处理)

- 层次化任务建模
- 增强型上下文管理



## Foundation Models (基座模型)

- ~~专属物理推理模型 (Xiwu-BESIII)~~
- 智能体动态路由与选择



## Tool Usage (工具调用)

- 通过 Qionwu (HaiDDF) 实现领域专属工具深度集成



## Code Generation (代码生成)

- 突破性的 BOSS 代码映射技术
- 领域特定语言 (DSL) 执行



Data Perception (数据感知): 增强的 PDF 视觉感知能力; 粒子物理拼写变体与特殊符号感知。



# Qionwu (HaiDDF) : 打破信息孤岛的分布式底座

将 LLM、科学工具、记忆和数据转化为可访问资源，将动态 Agent Workers 连接到超算集群。



## FAIR Principles

- Findable (可发现)
- Accessible (可访问)
- Interoperable (可互操作)
- Reusable (可复用)

# BOSS Code Worker: The Intelligent Executor

## From Chatbot to Expert Assistant.

### Functional Highlights

- Runs directly on High-Performance Computing (HPC) Clusters.
- Perceives complex experimental data.
- Executes multi-step cluster operations autonomously.

Run Bash  
Command



Executor



Message  
Interface



BOSS  
Code Worker  
(V2)



I/O Tools

View File

Write File

Inspect  
Environment



Git Tools

Show  
Difference

Undo  
Commit

The BOSS  
Code worker

DDF

a core tool

The Dr.Sai Agent

## 赋予系统“时间感知”：OpenClaw 与排程引擎

核心突破——“让系统学会等待”。

The screenshot displays the OpenClaw interface, which is designed for interacting with AI models. The main window is titled "Blueprint for Physics Intelligence and AI Infrastructure". The interface is divided into three main sections:

- 来源 (Sources):** Located on the left, it allows users to add new sources from the web or Fast Research. A list of sources is shown, including "未来规划-1页.pdf".
- 对话 (Chat):** The central area shows a document titled "Blueprint for Physics Intelligence and AI Infrastructure" with a summary of its content. Below the document, there is a chat interface with a "保存到笔记" (Save to notes) button and a "开始输入..." (Start input...) field.
- Studio:** Located on the right, it provides various tools and features for working with the document, including "音频概览" (Audio overview), "视频概览" (Video overview), "报告" (Reports), "测验" (Quizzes), "数据表格" (Data tables), "演示文稿" (Presentations), "思维导图" (Mind maps), "闪卡" (Flashcards), and "信息图" (Infographics).

The document content in the chat area reads:

这份资料概述了面向2030年的**基础物理与科学突破战略规划**，重点打造由数据、模型、智能体和基础设施构成的全方位体系。通过完善名为**Mamba、Daisy、Domas和Dora**的“3D”数据软件，该计划旨在建成**10PB级**的AI就绪型科学数据库。在技术层面，核心任务包括研发能理解**物理规律**的科学大模型，并推动**物理分析智能体**从原型设计走向实际应用。此外，通过优化**计算基础设施与领域工具链**，实现不同大科学装置间的**互联互通与资源共享**。整个蓝图旨在利用**人工智能**深度赋能科研流程，最终支撑**重大科学发现**的产生。





# Open Dr.Sai科研智能体框架 (一种Harness)

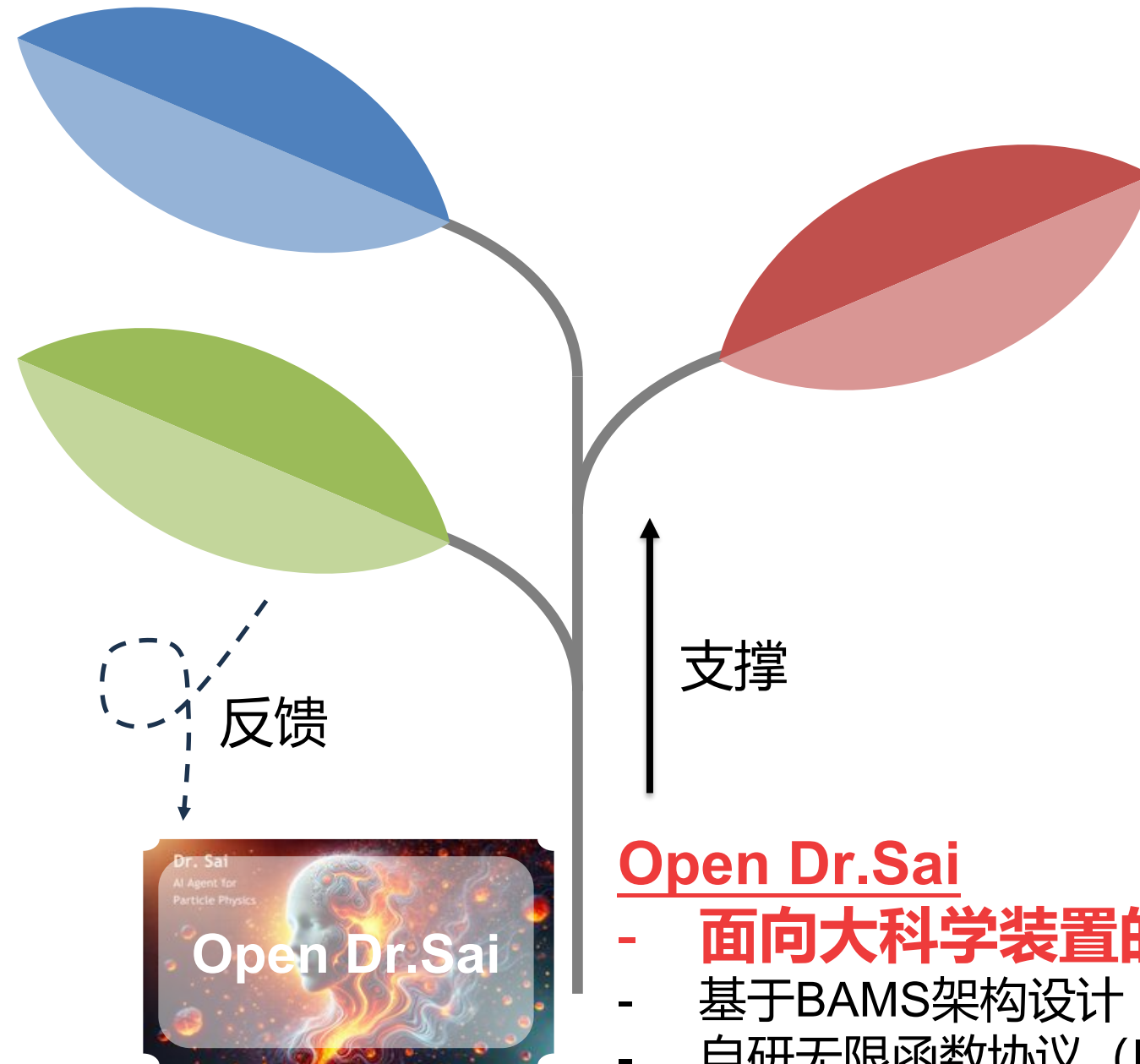
## 03 更多智能体

...

## 01 Dr.Sai BESIII

### 物理分析智能体

- 增强粒子物理能力
- 感知BESIII数据
- 装配BOSS工具
- 自动化完成复杂物理分析任务, 助力新粒子发现



## 02 DORA Agent

### 光源数据智能体

- 增强数据理解能力
- 感知光源中子源数据
- 装配谱学、成像、衍散射多方法学工具
- 自动科学数据加工、融合、解释和供给

## Open Dr.Sai

### - 面向大科学装置的科研多智能体框架

- 基于BAMS架构设计 (Brain-Actuators-Memory-Sensors)
- 自研无限函数协议 (IF), 支持MCP、A2A、AIP
- 多智能体协同和强可扩展性, 支持AutoGen、磐石、OpenClaw
- 完善的身份认证和细粒度的大装置权限认证

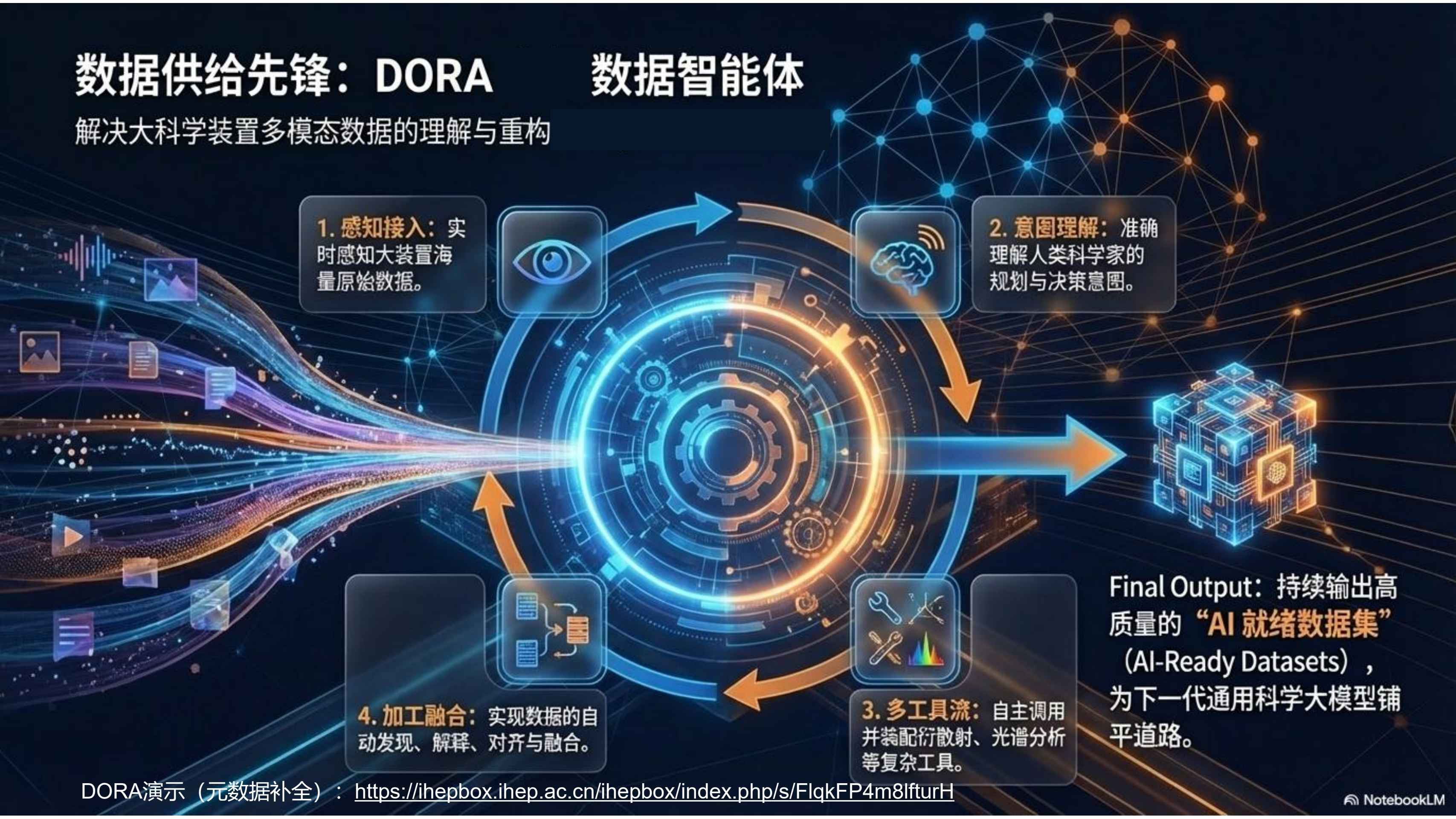
《开源仓库》: <https://github.com/hepai-lab/drsai>

《详细文档》: <https://docs-drsai.ihep.ac.cn/>

《智能体构建实用指南》: [https://note.ihep.ac.cn/s/\\_vSnuf5Uh](https://note.ihep.ac.cn/s/_vSnuf5Uh)

# 数据供给先锋：DORA 数据智能体

解决大科学装置多模态数据的理解与重构



1. 感知接入：实时感知大装置海量原始数据。



2. 意图理解：准确理解人类科学家的规划与决策意图。



3. 多工具流：自主调用并装配衍散射、光谱分析等复杂工具。

4. 加工融合：实现数据的自动发现、解释、对齐与融合。



Final Output: 持续输出高质量的“AI就绪数据集” (AI-Ready Datasets), 为下一代通用科学大模型铺平道路。

# 实战验证：BESIII 物理分析智能体的卓越效能

证明智能体具备完全自主驱动复杂多步物理分析研究的能力。

**~40 个子步骤 (Substeps)**

每次物理分析任务的极高复杂度。

**100 次高强度采样 (Samplings)**

AI幻觉

跨越 10 个不同的衰变道 (Decay Channels) 进行严格验证。

**>90% 成功率 (Success Rate)**

端到端自动化执行的极高可靠性。

**10倍 效率提升 (10x Efficiency)**

彻底打破传统人工分析的耗时瓶颈。

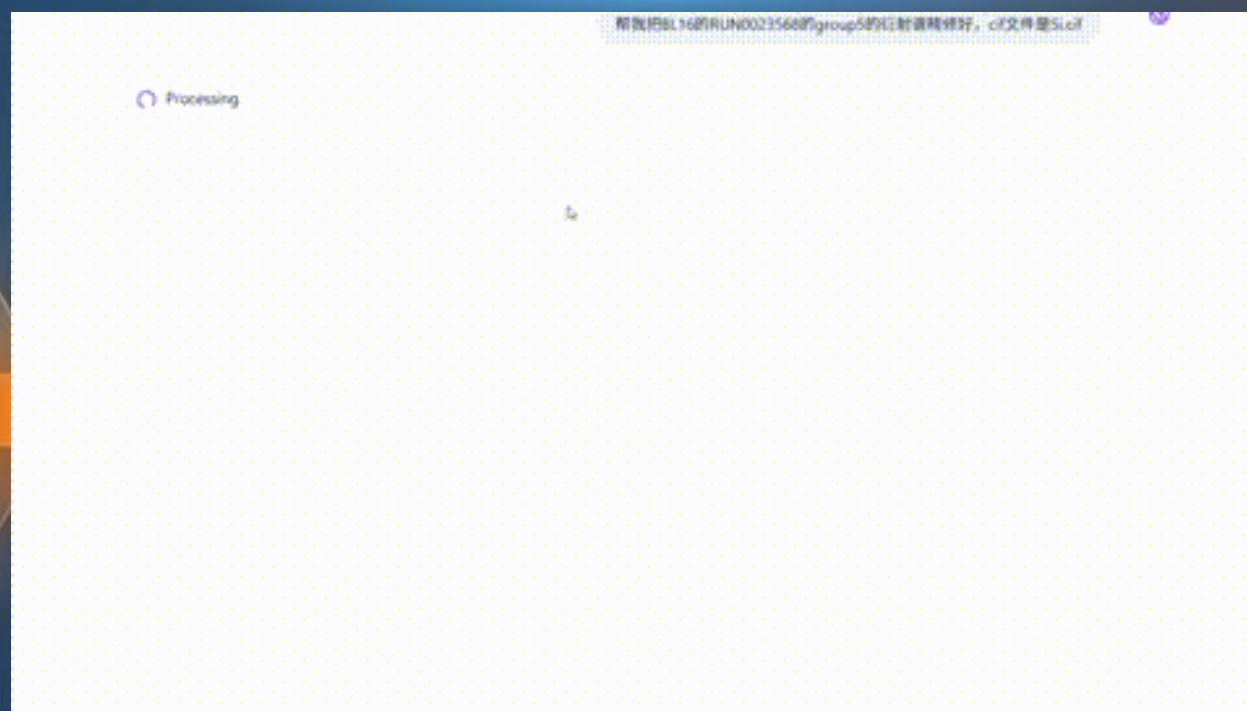
让大型语言模型真正掌握通用的物理分析过程，从底层释放科研生产力。

# 攻克复杂精修: Rongzai 衍射智能体

面向中国散裂中子源 (CSNS) 的无人化数据解析专家 (Target: 2026年1月)



**复杂样本挑战:** 粉末中子衍射数据包含极高的噪声与重叠峰, 传统精修需要专家耗费数周时间手动调参。



**Unattended Operation (无人干预):** 智能体全自动接管精修流程。

**Target Materials:** 从标准样品 (如硅 Si) 到具有极端复杂晶体结构的样品 (如沸石)。

**Outcome:** 提供高精度、物理一致的精修结果, 大幅缩短材料研发周期。

# 战略路线图：从原型走向率先实用

打造面向 2030 年基础物理与科学突破的  
10PB 级“3D”数据软件与智能体生态体系。

## Phase 1: 概念与原型 (Prototype System)

Open Dr.Sai v2.0 发布，验证复杂环境下的 BAMS 架构与多智能体协同。

## Phase 2: 领域泛化 (General Physical Analysis)

覆盖更多衰变道，实现对粒子物理、光源数据的通用理解 (2025)。

## Phase 3: 率先实用 (Practical Real-World Application)

跨大科学装置的广泛部署，成为支撑重大科学发现的互联互通平台 (2026+)。

探索开源生态与共建：<https://github.com/hepai-lab/drsai>

# 基于OpenDRSAI的个人助手



所内专属独立IP



基于容器的环境隔离

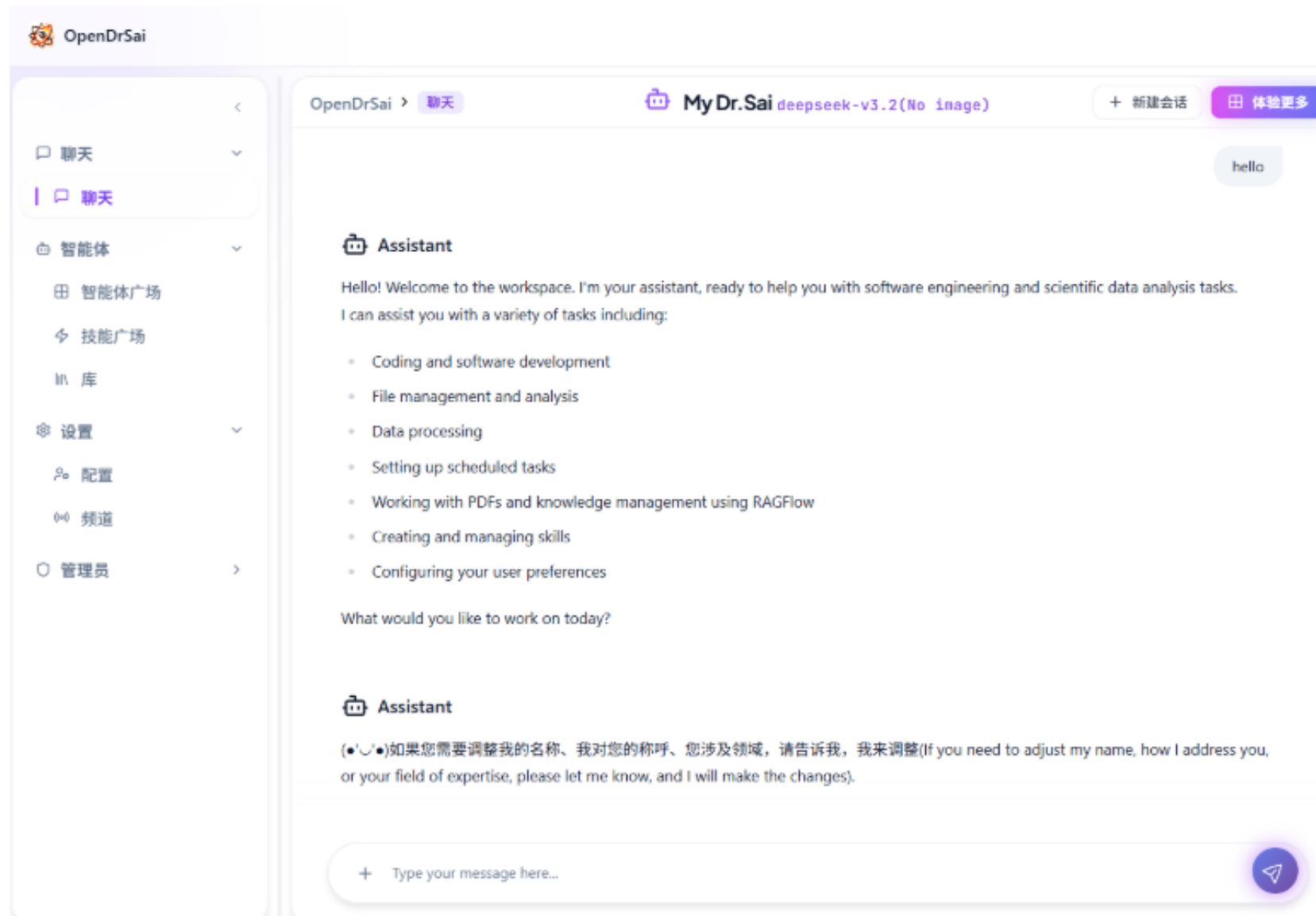


数据持久化

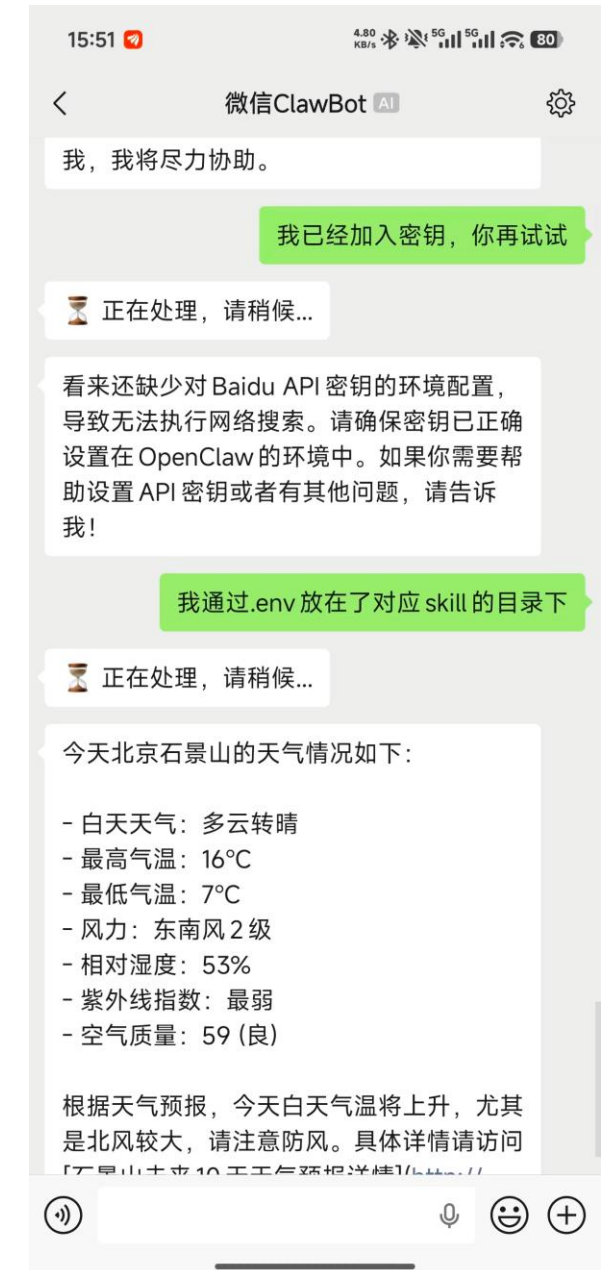
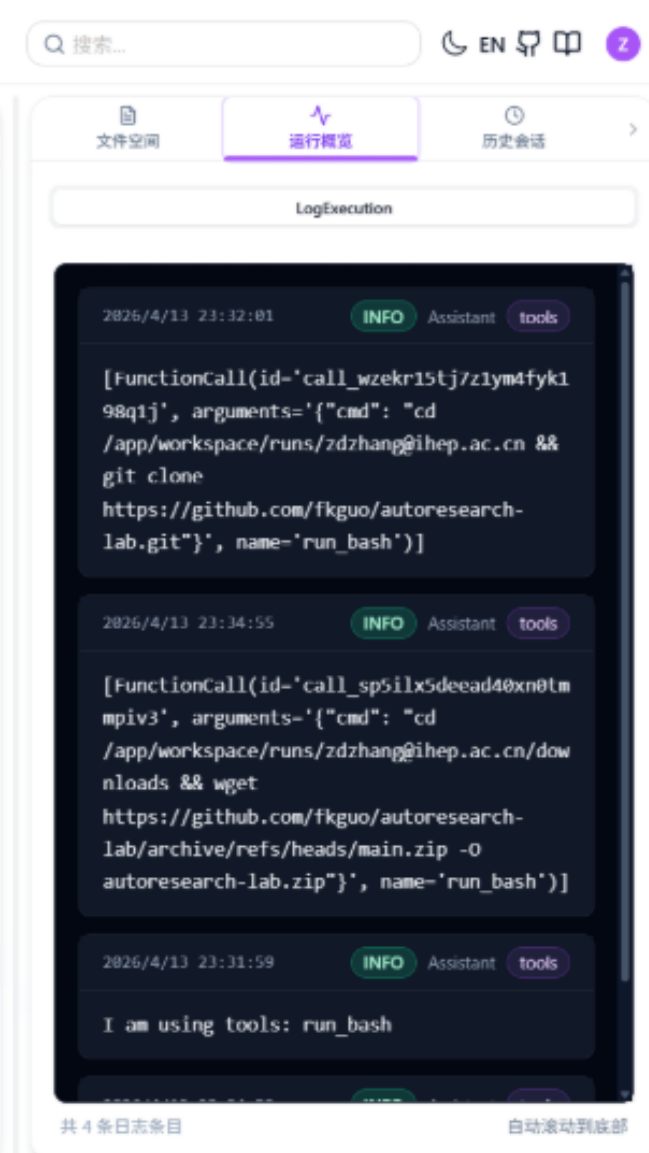


本地模型和云端模型无缝切换

http://10.5.8.xx:18789/



OpenDRSAI界面

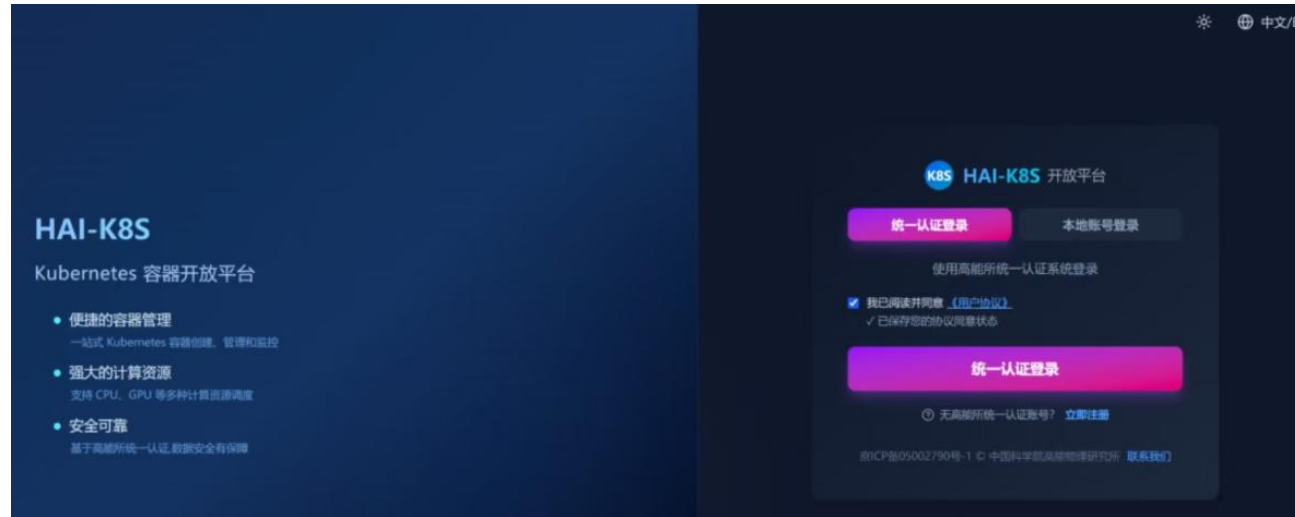


接入微信等app

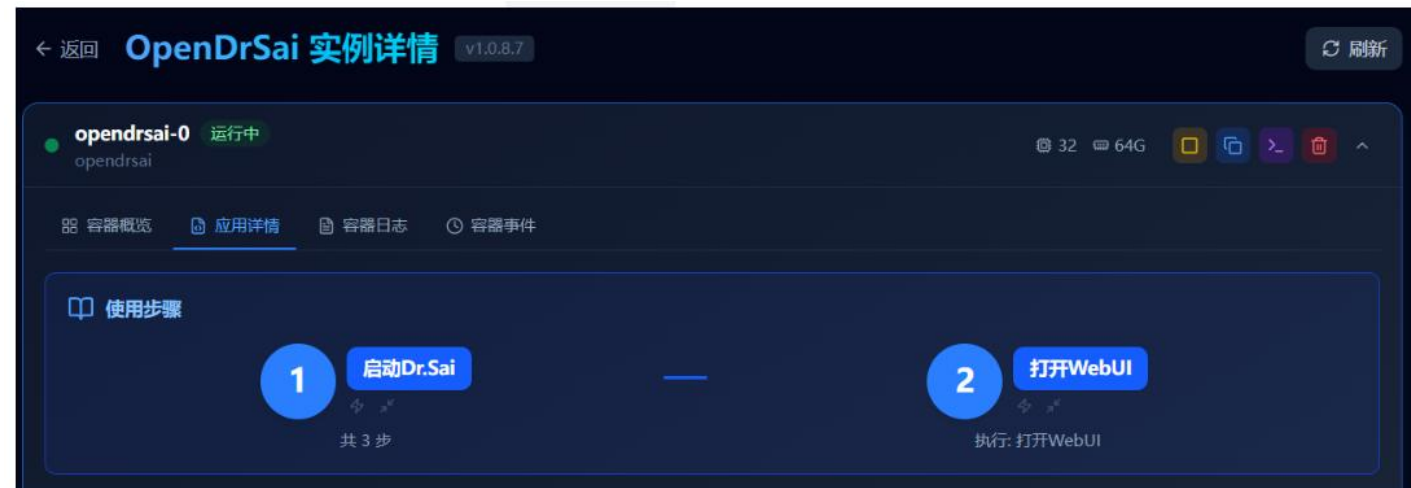
# OpenDrSai使用方法



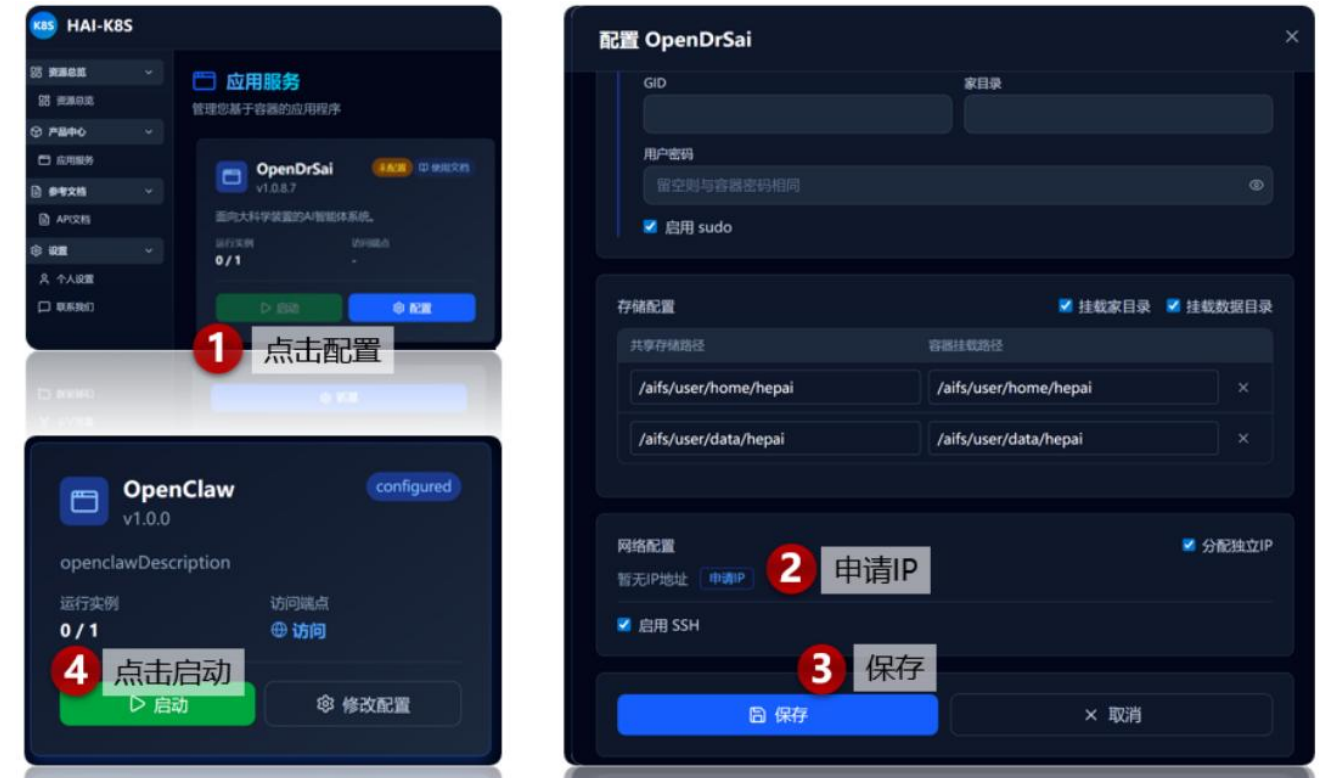
第一步：登录HAI-K8S开放平台：<https://k8s-ai.ihep.ac.cn>



第三步：启动Dr.Sai



第二步：配置应用和启动容器



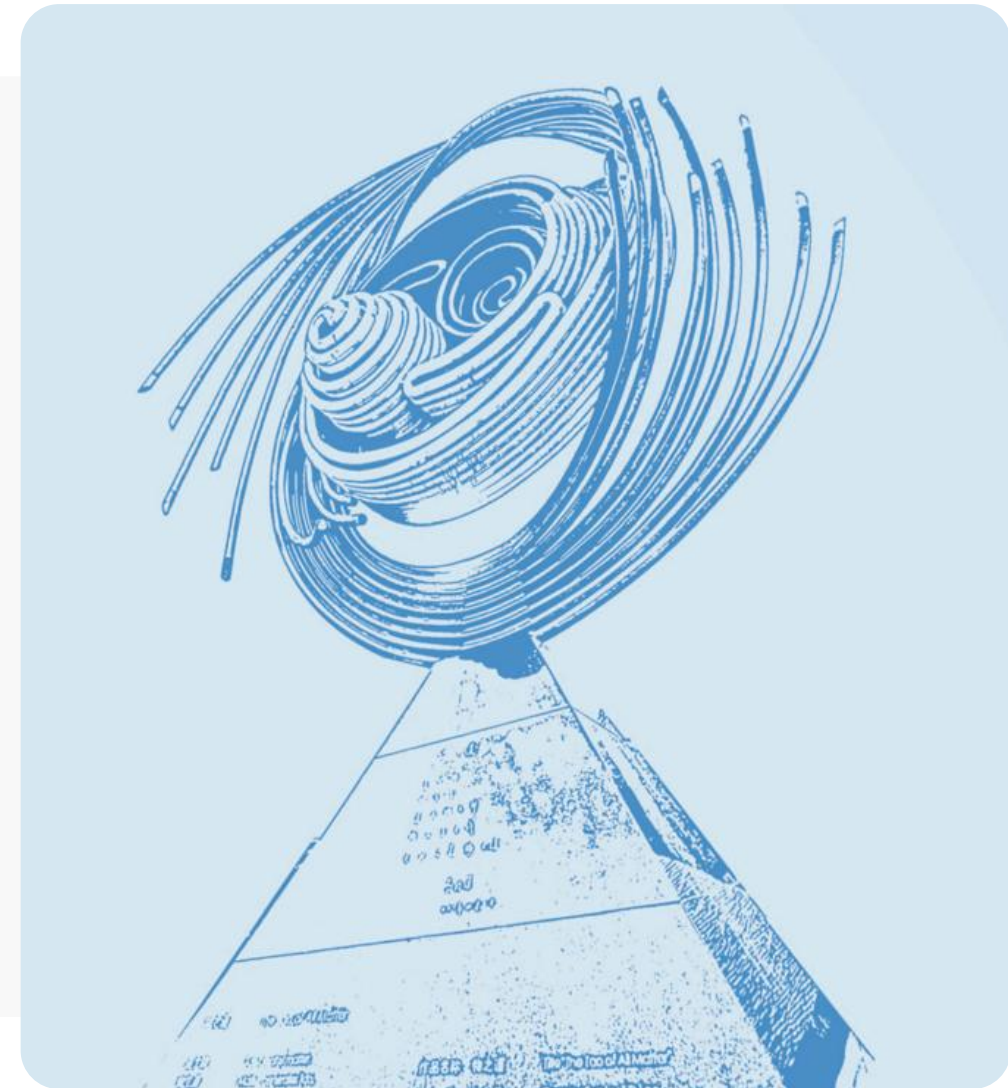
第四步：打开界面开始使用“My Dr.Sai”！



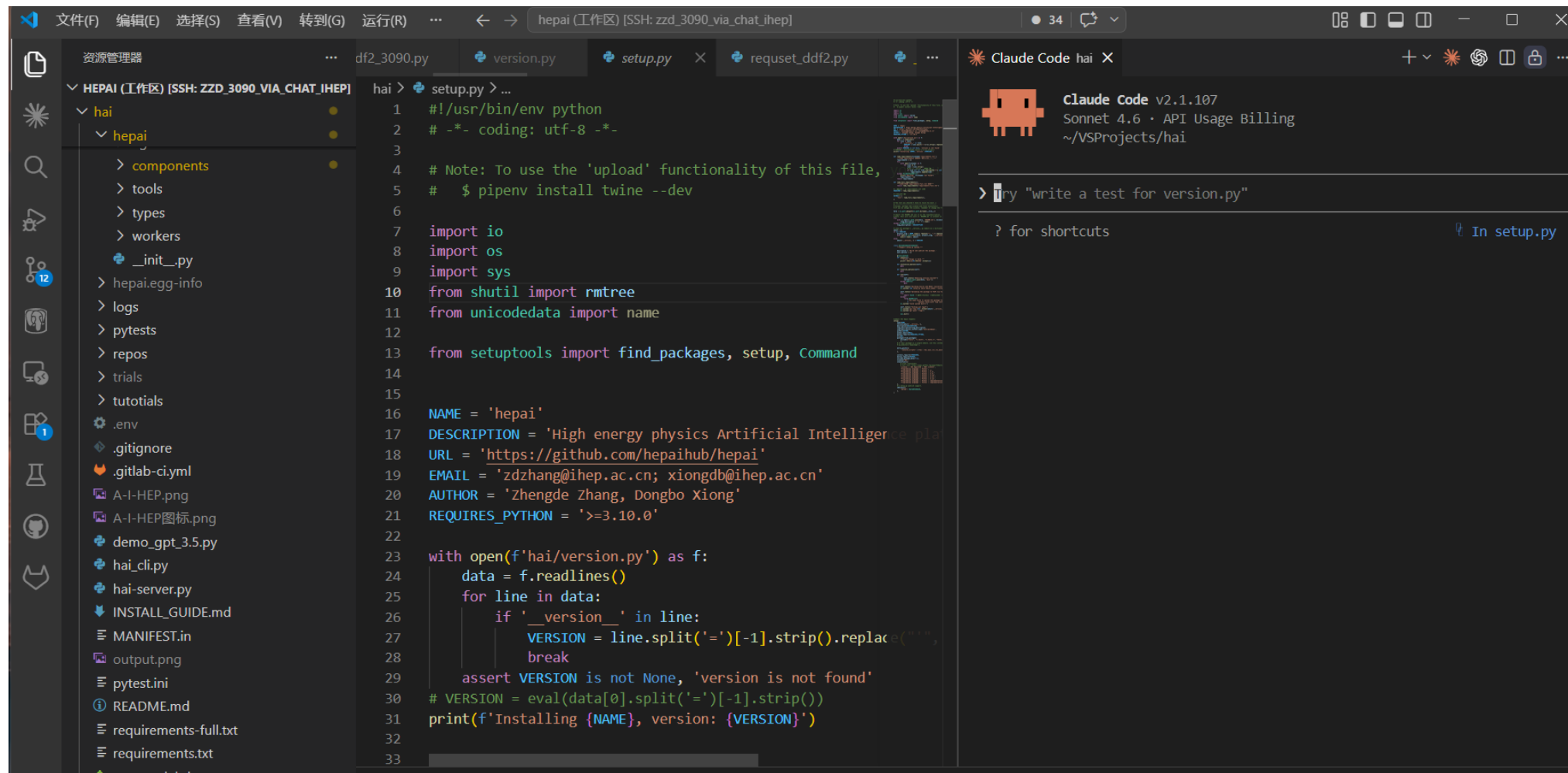
【详细教程】一键拥有您的“赛博士”智能助手：<https://note.ihep.ac.cn/s/3QMhFrs8n>

Part X

# Backup



# Claude Code的运行环境



| 工具类别 | 工具名称         | 并发安全 | 主要功能       |
|------|--------------|------|------------|
| 文件操作 | Read         | ✓    | 文件内容读取     |
|      | Write        | ✗    | 文件内容写入     |
|      | Edit         | ✗    | 文件内容编辑     |
| 搜索发现 | MultiEdit    | ✗    | 批量文件编辑     |
|      | Glob         | ✓    | 文件模式匹配     |
|      | Grep         | ✓    | 内容正则搜索     |
| 任务管理 | LS           | ✓    | 目录结构列举     |
|      | TodoRead     | ✓    | 任务列表查看     |
|      | TodoWrite    | ✗    | 任务列表更新     |
| 系统执行 | Task         | ✓    | SubAgent启动 |
|      | Bash         | ✗    | 命令行执行      |
| 网络交互 | WebFetch     | ✓    | 网页内容获取     |
|      | WebSearch    | ✓    | 搜索引擎查询     |
| 特殊功能 | NotebookRead | ✓    | Jupyter读取  |
|      | NotebookEdit | ✗    | Jupyter编辑  |

# 领域专属先锋智能体矩阵



## Dr.Sai BESIII (粒子物理智能体)

**Target Facility:** BESIII 探测器

**Core Capability:** 感知高能物理数据，装配 BOSS 工具链。

**Mission:** 自动化完成复杂物理分析任务，助力新粒子发现。



## DORA Agent (光源数据智能体)

**Target Facility:** 光源与中子源 (HEPS/CSNS)

**Release:** 2025.12

**Core Capability:** 感知光源中子源数据，装配谱学、成像、衍散射多方法学工具。

**Mission:** 自动处理科学数据（对齐、标注、融合），输出高质量 AI 就绪数据集。



## Rongzai Agent (衍射精修智能体)

**Target Facility:** CSNS (中国散裂中子源)

**Release:** 2026.01

**Core Capability:** 复杂中子衍射数据的智能化处理。

**Mission:** 在无人干预下，自动完成标准样品 (Si) 及复杂样品 (沸石) 的粉末中子衍射精修。