



Institute of High Energy Physics, CAS

National HEP Data Center

AI@IHEP-CC

人工智能赋能高能物理的实践与展望



Zhengde Zhang (张正德)

Computing Center, Institute of High Energy Physics, CAS

2025.08.26



目 录

1

数据体系

2

AI模型和应用

Chinese Academy of Sciences

3

大模型智能体

4

HepAI平台

5

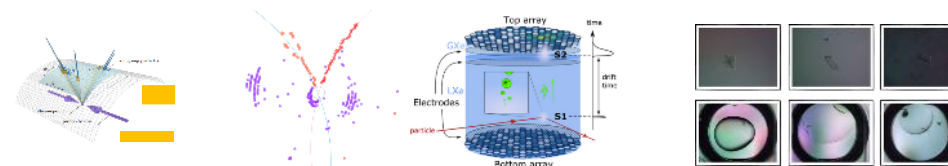
总结与展望

发展面向高能物理的AI总体战略



The 14th Five-Year Plan of the IHEP:

- Advanced computing technologies including AI, QC make more powerful modeling and simulation possible, which are indispensable means for major breakthroughs in HEP.
- It is necessary to strengthen the application of ML in **real-time processing, simulation, reconstruction, analysis**, etc., to enhance the ability of high energy physics exploration and new discoveries.



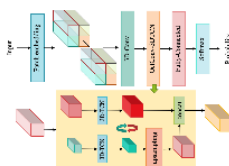
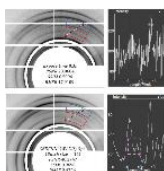
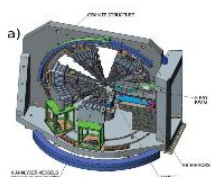
Our Exploration experiences:

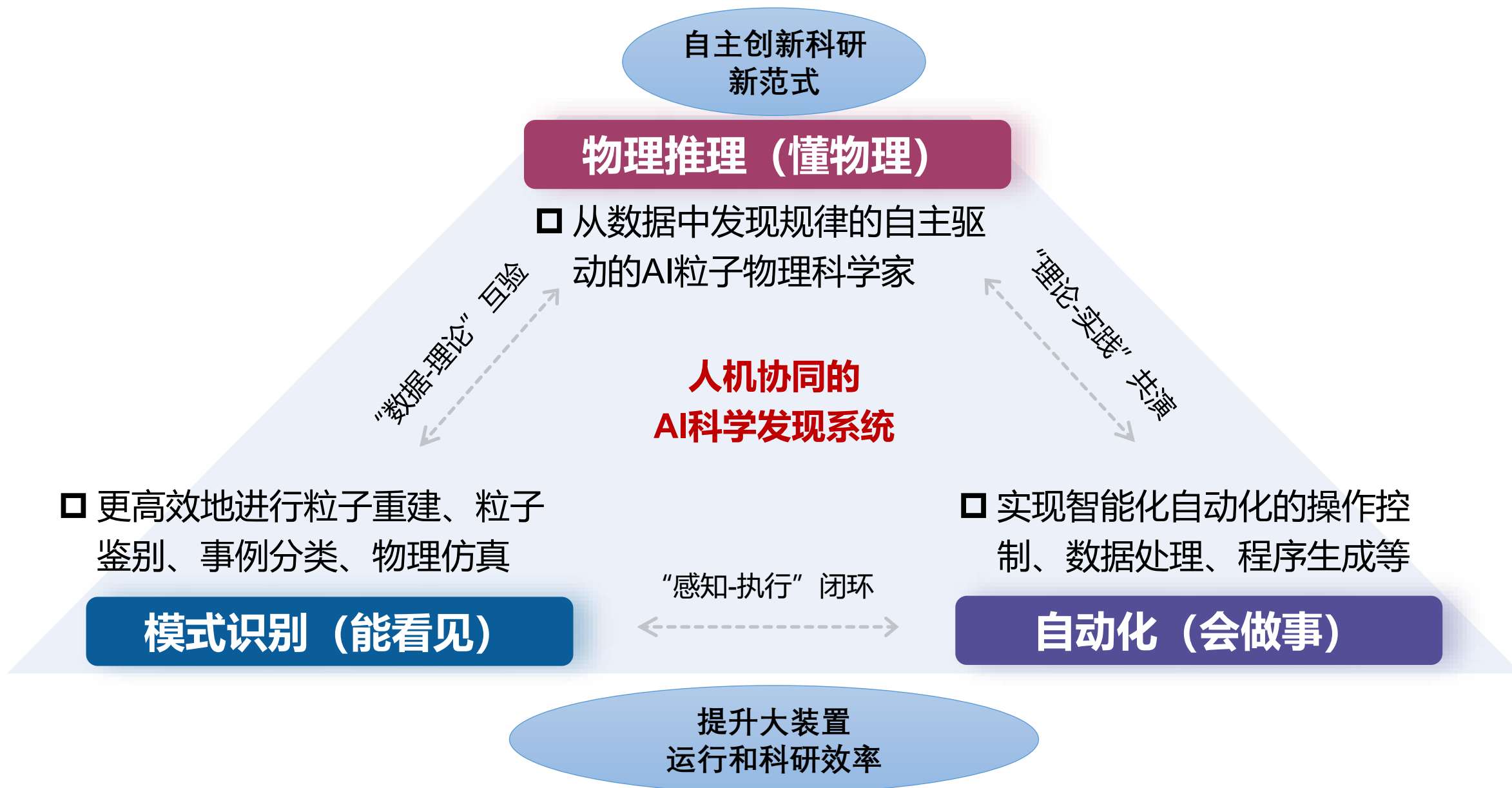
- 十四五:
- 新兴前沿方向和未来技术



- 十五五:
- 大科学装置关键技术

- Prioritizing the development of **AI-Ready scientific datasets** is of utmost importance.
- Insist on the concurrent development of **software frameworks** and **specific AI applications**.
- Start investigating the **large models** to propel new discoveries of HEP.

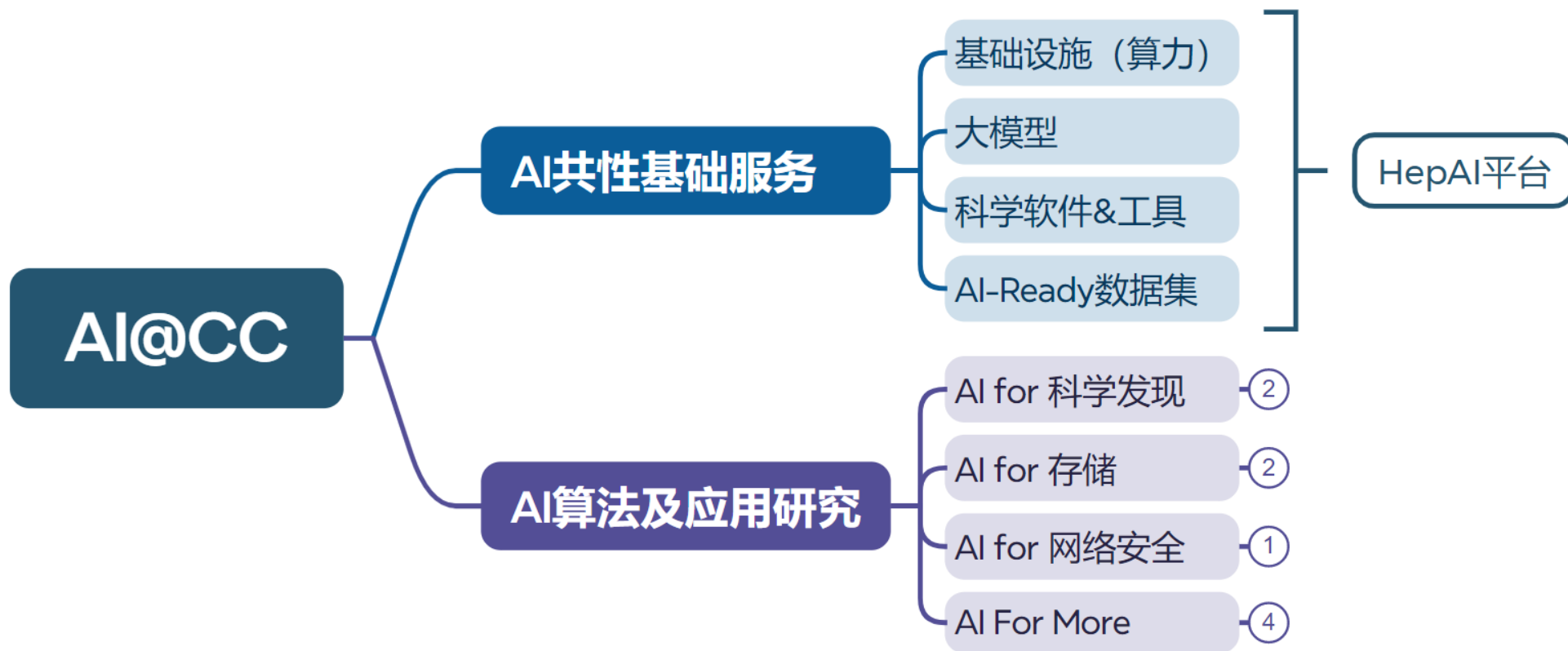


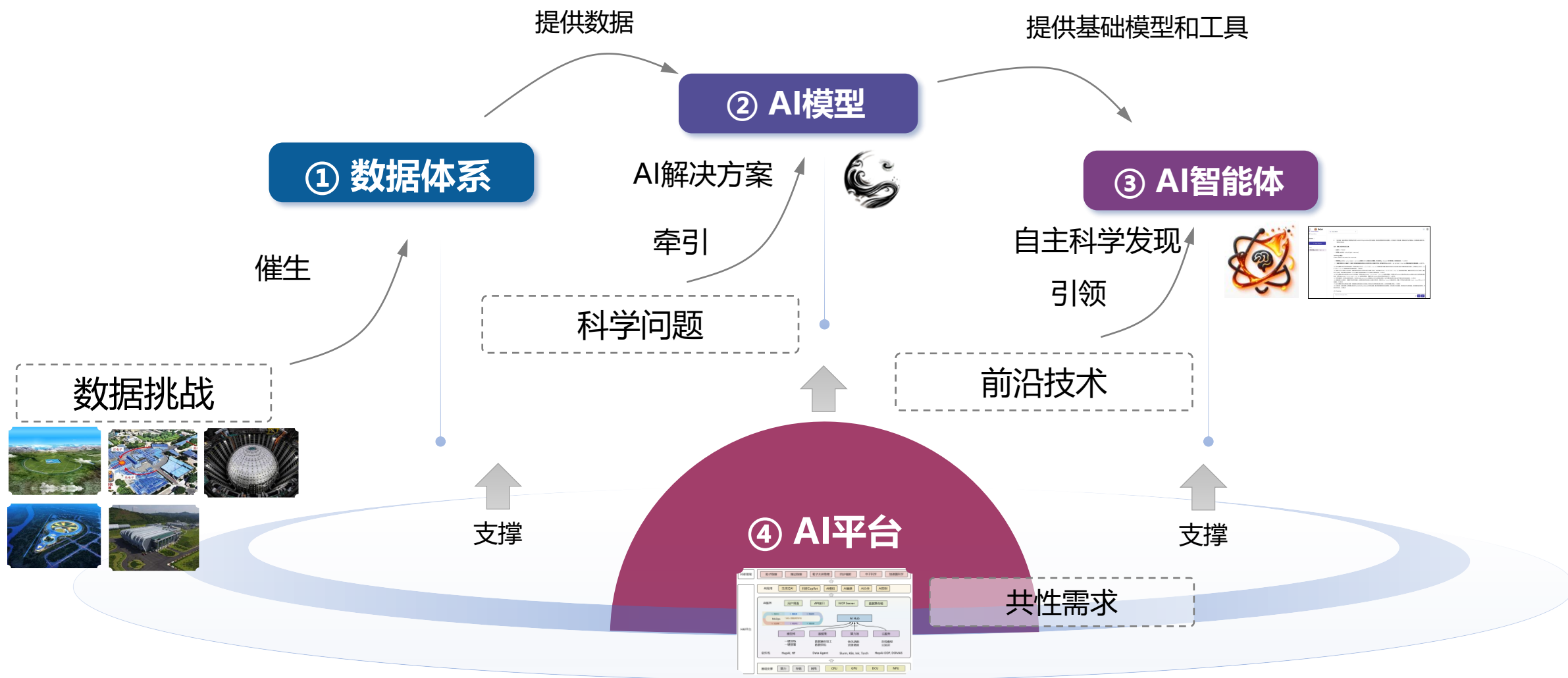


高能所计算中心AI总体规划



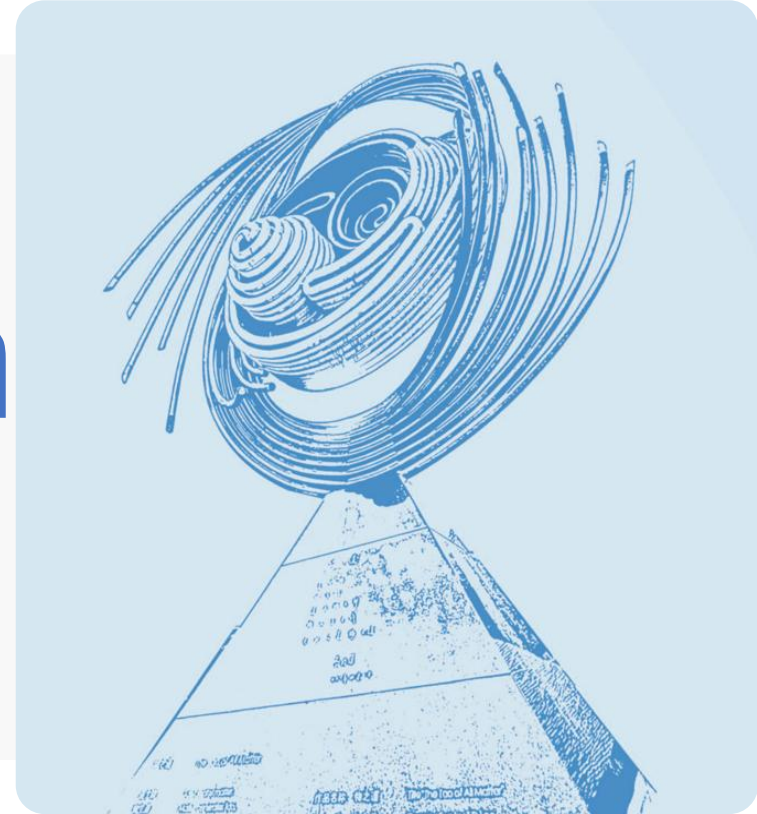
- 人工智能@计算中心的定位：**服务+科研融合**
- **AI共性基础服务**（面向领域科研活动）：**算力基础设施、大模型、工具、数据集**→**高能物理AI平台**
- **AI算法及应用研究**（面向信息技术发展）：**科学发现、存储、安全、etc.**





Part 1

Data System



The AI-Ready Data: A Must



- AI三要素：数据、算法和算力
- 目前大装置科学数据状态：统一的格式、统一的存储，但大部分是非AI-Ready的
- 数据类型：实验数据、模拟数据、语料数据、装置运行数据

Data tools/ Data agents



Data cleaning

Metadata Completion

Data content explanation

Data deep search/discovery

Professional data processing

Data Platform



什么是AI-Ready数据集?



AI-Ready数据集：能够被**高效、安全、可复现**地用于**训练、评估和部署**AI/ML系统的数据集合。

核心特征：

① 任务适配

- 与目标场景/任务强相关
- 覆盖面和代表性

② 高质量与一致性

- 准确、完整、一致
- 去重、噪声可控

③ 本体与标注

- 高质量标签
- 层次/本体映射关系，标注审计

④ 工程可用

- 机器可读、标准格式
- 合理分片/分桶、可流式、可并行

⑤ 评估与可复用

- 严格训练/测试/验证划分
- 基准集与指标明确

⑥ 元数据丰富性

- 采集方式、时间、设备/系统
- 上下文、版本信息

⑦ 数据偏差控制

- 采样偏差、标签偏差、历史偏差

⑧ 可获取与可复用

- 稳定访问接口(API/数据湖/特征库)
- 文档与示例

⑨ 合规和伦理

- 许可与使用权
- 隐私保护、最小PII

⑩ 安全与可靠

- 加密（传输/静态）
- 最小权限、密钥管理



大装置数据体系是大科学装置中数据的采集、存储、处理、分析和使用的整体框架和机制。它涵盖数据从生成到最终使用的整个生命周期，并确保数据的质量、完整性、安全性和可用性。

数据策略与标准

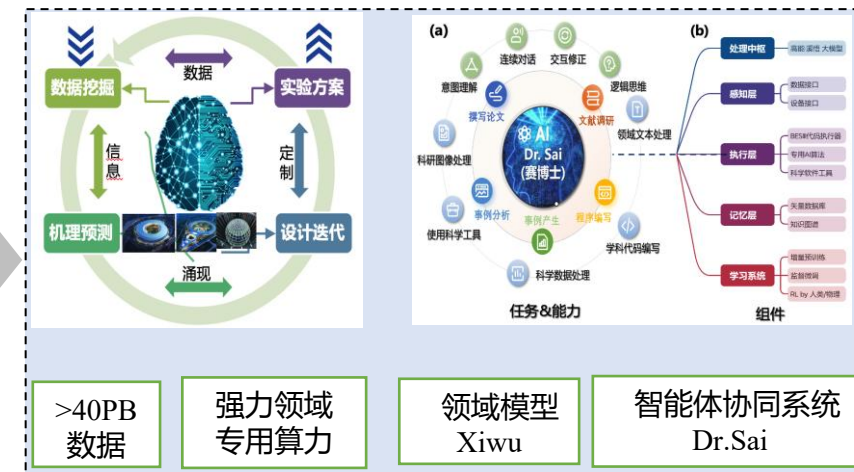
- **数据模型**
 - 表征数据
 - 元数据
- **数据标准**
 - 数据格式
 - 元数据包
- **数据策略**
 - 保存策略
 - 使用机制

数据软件与工具

- **采集与处理Mamba**
 - 实验设施运行控制
 - 实验过程控制
- **存储与管理DOMAS**
 - 大数据存储
 - 元数据管理
- **分析与应用Daisy**
 - 科学数据分析软件
 - 数据可视化
- **AI-Ready数据构建HEPAI**
 - 数据标注与增强
 - 多源数据融合

数据服务与供给

- **面向领域科学家**
 - AI和数据接口
- **面向AI研究人员**
 - 数据发现
 - 数据解释
- **面向AI机器人**
 - 数据管道
 - 机器可用



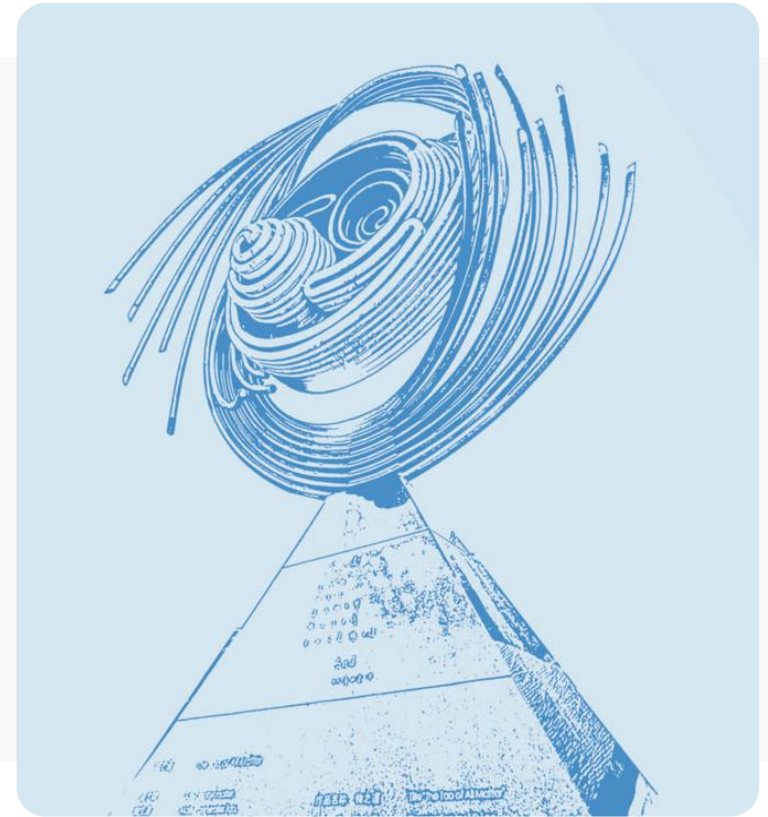
智慧型大科学装置

AI 驱动的科学发现系统

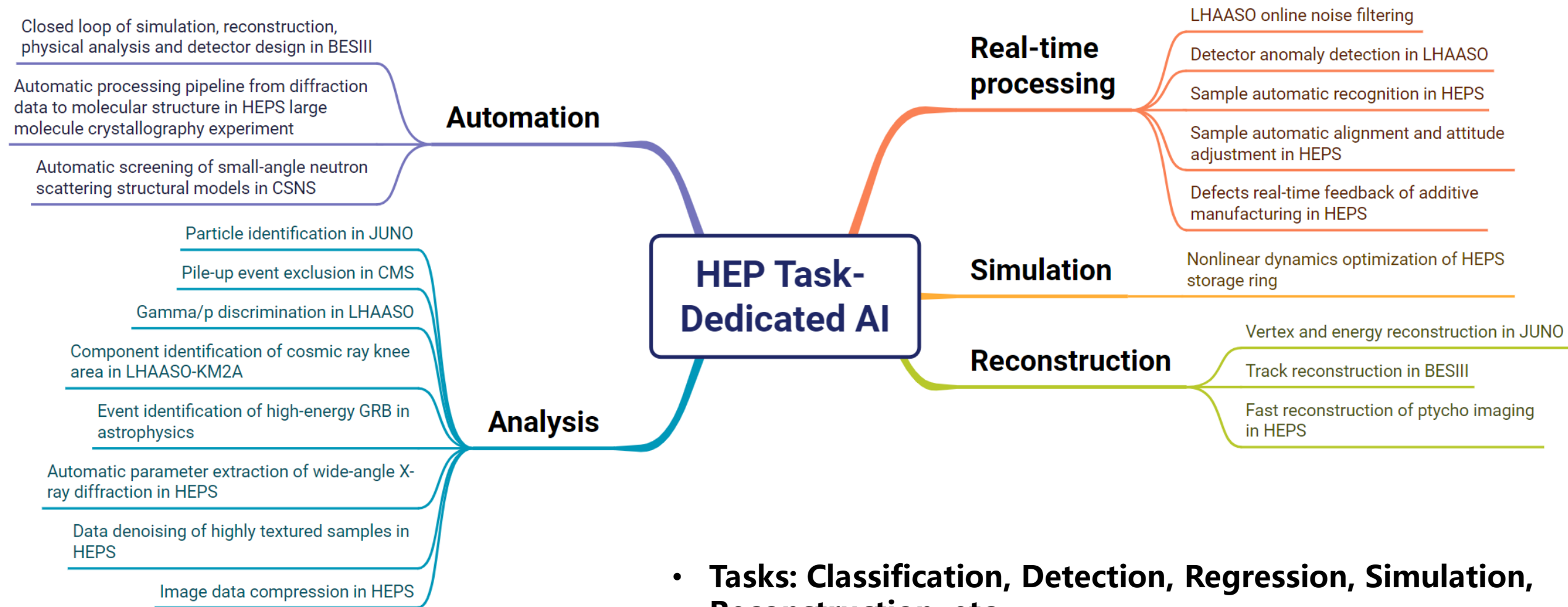


Part 2

AI Models & Application



Overview of HEP Task-dedicated AI

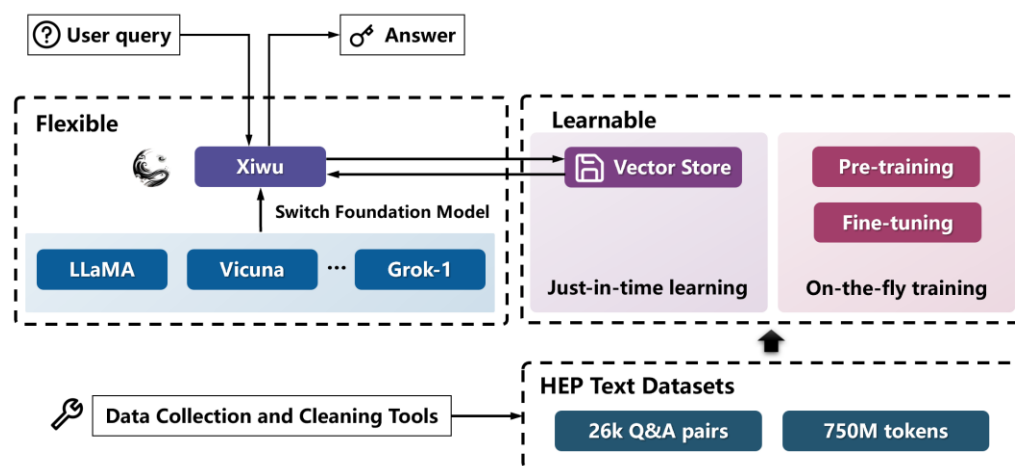


- **Tasks: Classification, Detection, Regression, Simulation, Reconstruction, etc.**
- **More than 20 applications**

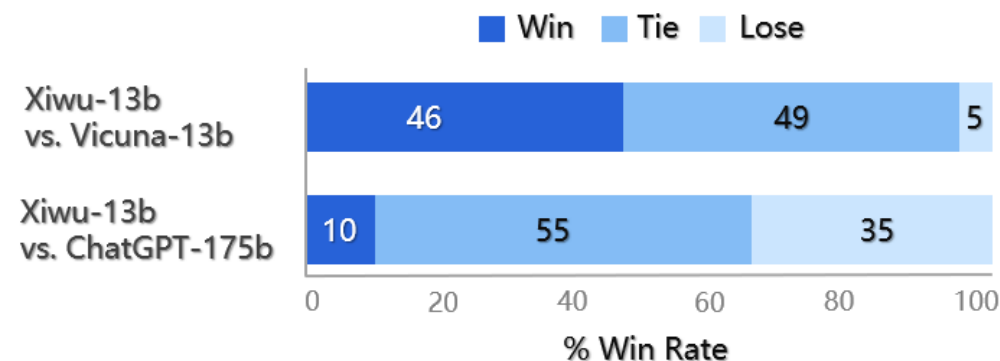


Brain **Xiwu(溪悟)** : A Basis Flexible and Learnable LLM for High Energy Physics ¹

- ❑ The **Brain** of Dr.Sai Agent.
- ❑ Currently, Xiwu is based on **DeepSeek-R1-Distill-Llama-70B**, Historically, LLaMA1~3.
- ❑ Secondary pre-training and fine-tuning techniques.
- ❑ Trained on GPU-A100 previously, now trained on DCU-K100 (Domestic GPU)
- ❑ Significantly better than the base model in HEP Q&A and internal code generation.
- ❑ New Model: **xLAM-2-32b-fc-r**, for function calling



Test Results



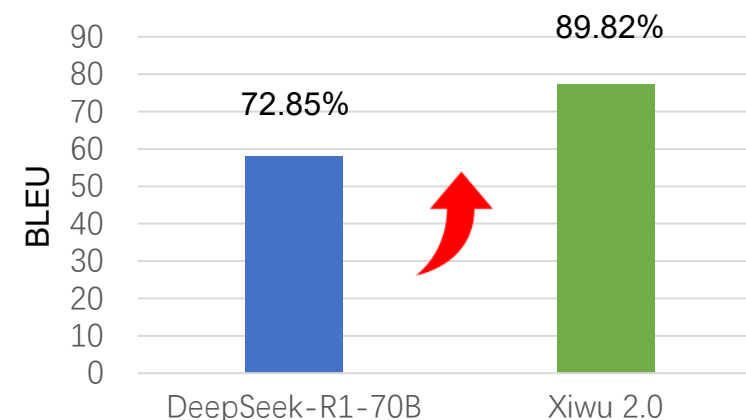
¹ [arXiv:2404.08001](https://arxiv.org/abs/2404.08001)



Development of Domain-specific LLM

Xiwu 2.0

- ❑ Based on **DeepSeek-R1-Distill-Llama-70B**.
- ❑ Fine-tuned on 39k Q&A pairs, including mixed samples from both HEP and common corpus datasets.
- ❑ Trained on DCU-K100-AI (Domestic GPU).



Comparison on HEP domain-specific dataset.

Domain-specific LLM for Dr.Sai Multi-Agent System

Dr.Sai-Host

- ❑ Built on DeepSeek-R1-Distill-Llama-8B and **xLAM-2-32b-fc-r**
- ❑ Finetuned using runtime data of Dr.Sai
- ❑ Optimized for LLM tool calling, specifically in expert agent selection.

Dr.Sai-Planner

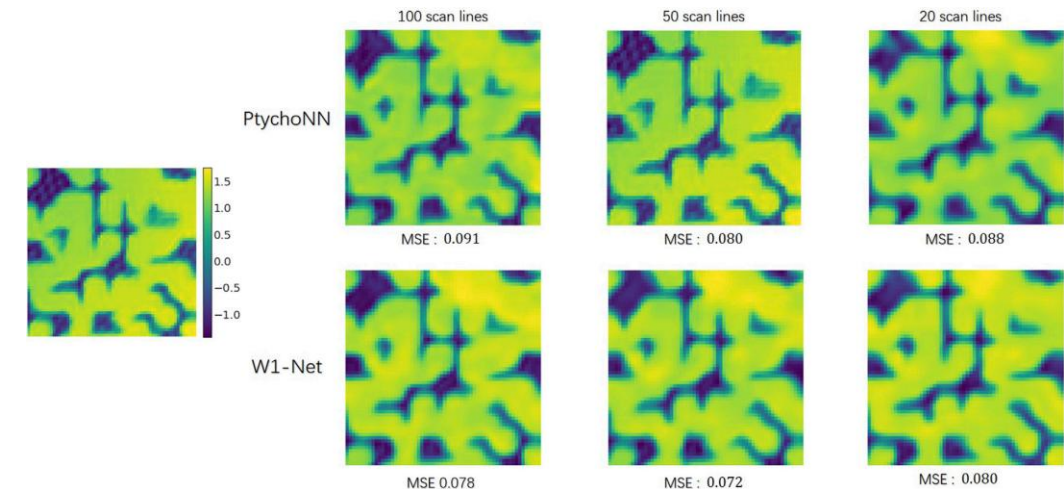
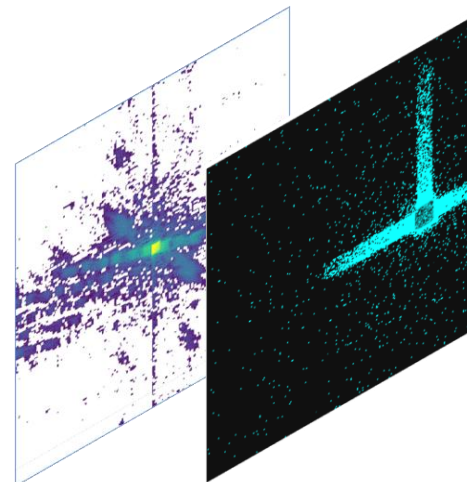
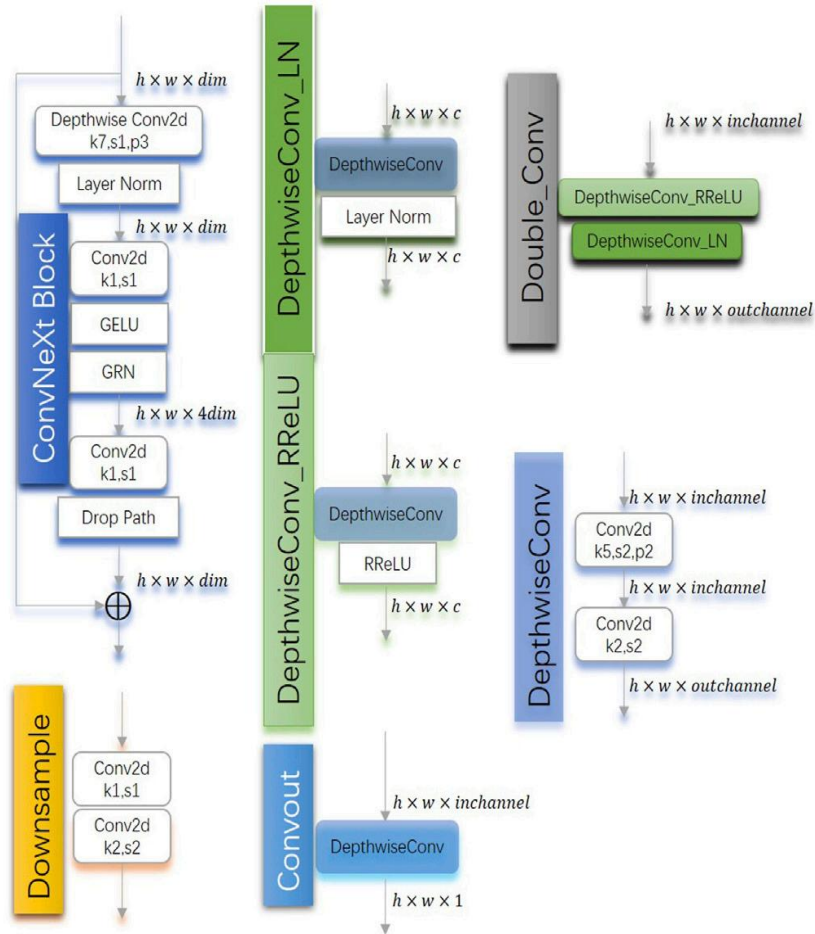
- ❑ Currently under development, fine-tuned with runtime data of Dr.Sai
- ❑ Specialized in planning particle physics data analysis workflows.

AI Model for Ptycho Imaging Reconstruction



Lei Wang, Guangcai Chang

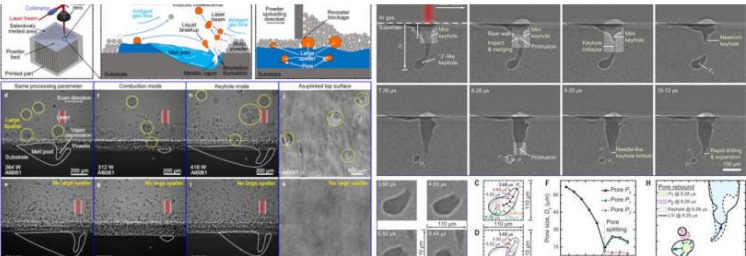
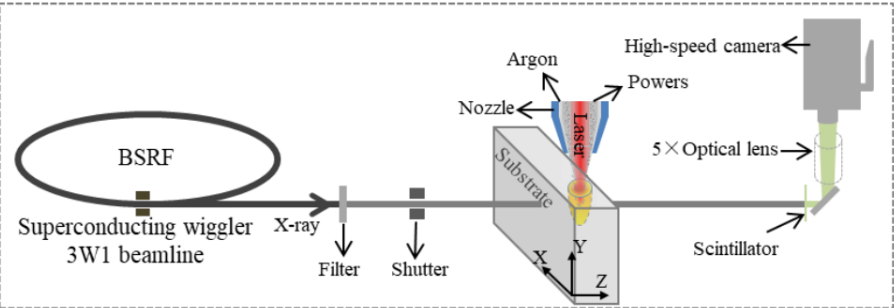
W1-Net, deep learning method for ptycho
80% less data volume, **100X faster** reconstruction speed
deploy on the **Jetson Nano** lightly
proposed by IHEP, 2024





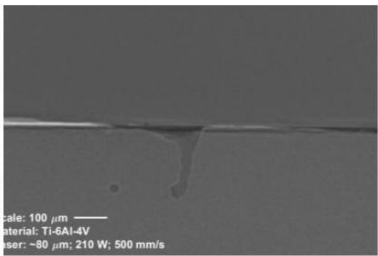
基于AI的同步辐射金属增材制造超快X射线成像的原位监测和工艺反馈

- 成像帧频: 80k/s (张兵兵)
- 单图: 1024*1024*12bit

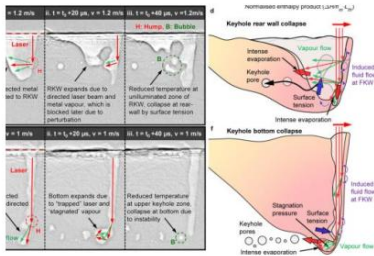


Nat. Commun. 13,1079, 2022

Phys. Rev. X 9, 021052 (2019)



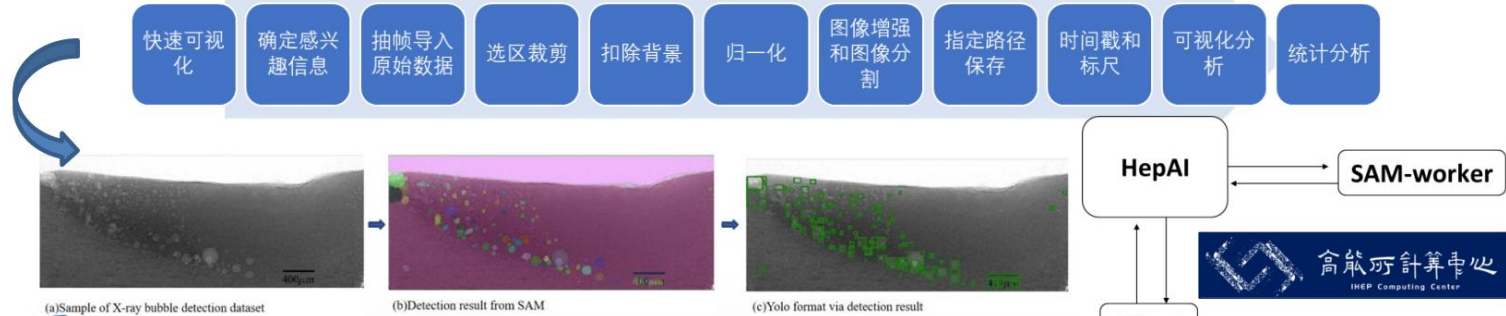
Nat. Commun.13,1170,2022



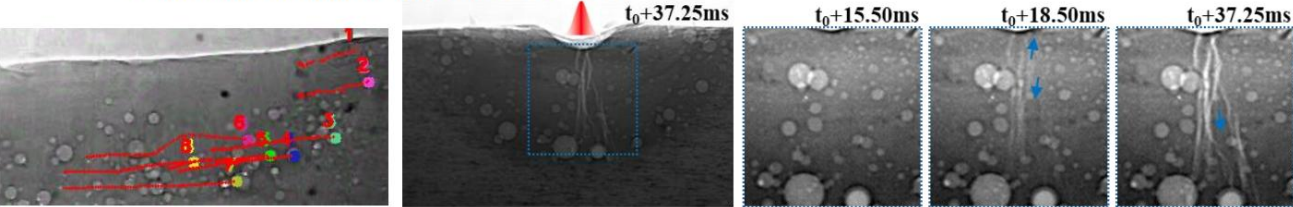
Science 370, 1080-1086 2020

基于BSRF的同步辐射超快X 射线成像的激光增材制造原位监测示意图。气孔形成机理及其动力学

匙孔波动动力学及气泡演变机制

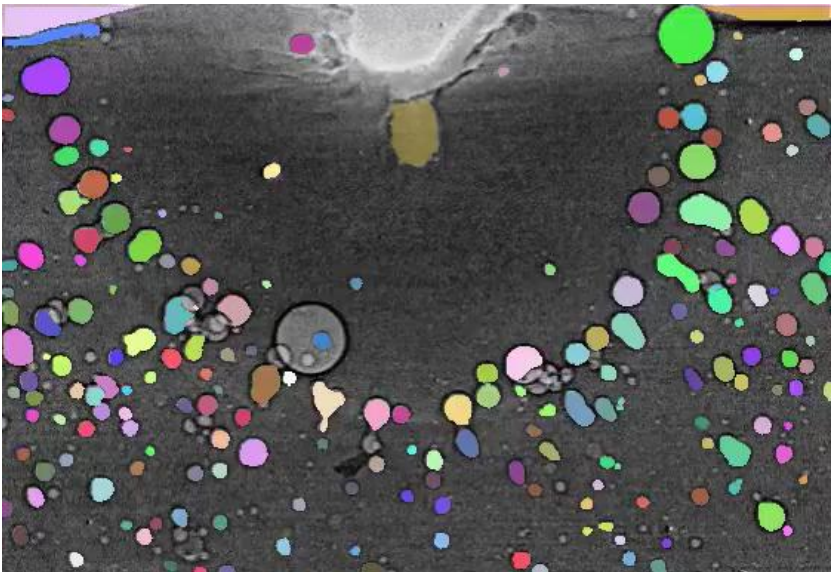


AI auto-annotation method based on SAM



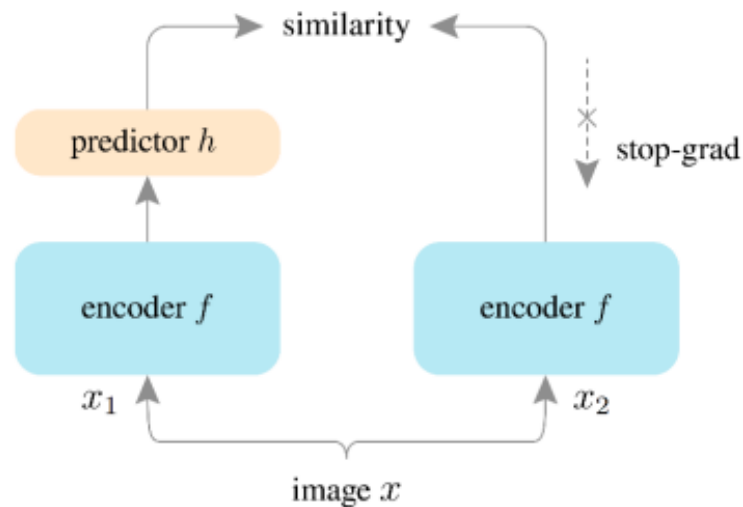
气泡消除率

气泡和裂纹的演化相关

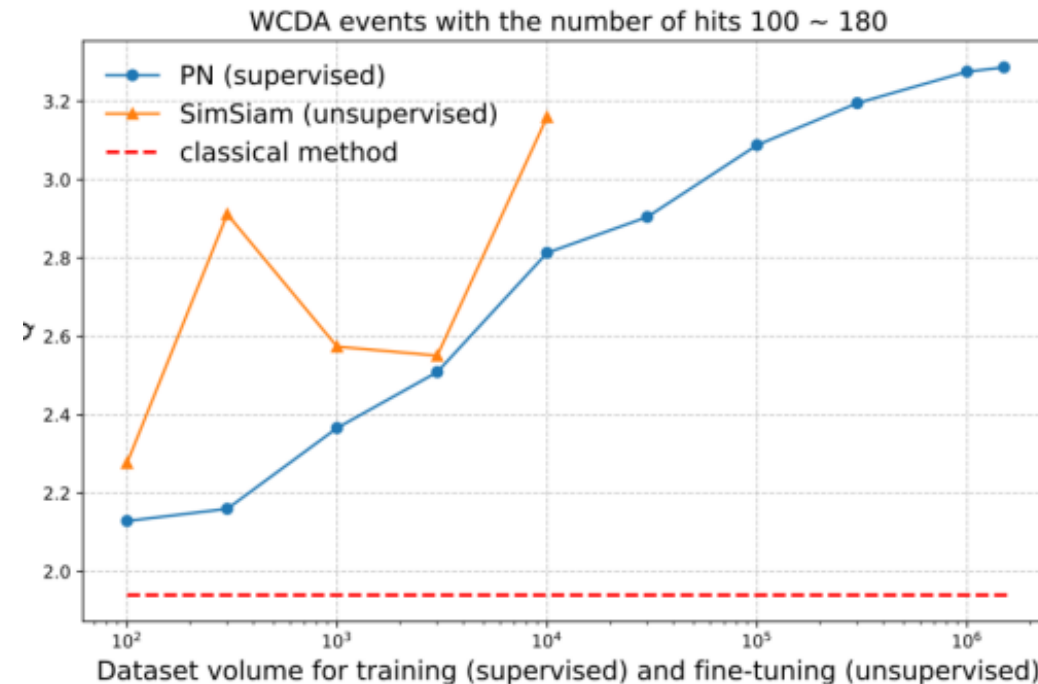


Self-Supervised SimSiam [1]:

After augmenting sample x_1 to x_2 , both pass through a shared encoder.
A predictor then pulls close feature similarity between original and augmented instances.



Pre-training encoders with 3×10^6 unlabeled samples, then fine-tuning with sparse labels
 $\Rightarrow$ outperforms supervised-only training.



Toward Foundation Model: CLPP



Zhengde ZHANG

CLPP

A foundation model trained by Contrast Language Particle Pre-training

Data

e.g.: HEP Particle Data (Jet)

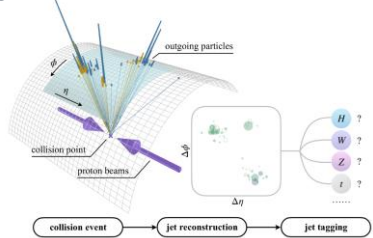


表 4.4 数据集特征定义

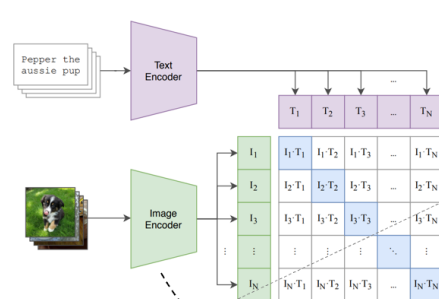
特征	定义
$\Delta\eta$	粒子与喷注轴之间的伪快度 η 的差值
$\Delta\phi$	粒子与喷注轴之间的方位角 ϕ 的差值
$\log p_T$	粒子横动量 p_T 的对数
$\log E$	粒子能量 E 的对数
$\log \frac{p_T}{E(jet)}$	粒子相对于喷流 p_T 的横动量对数
$\log \frac{E}{E(jet)}$	粒子能量相对于喷流能量的对数
ΔR	粒子与喷注轴之间的角分离 $(\Delta\eta)^2 + (\Delta\phi)^2$

Input: 7 features * n particles

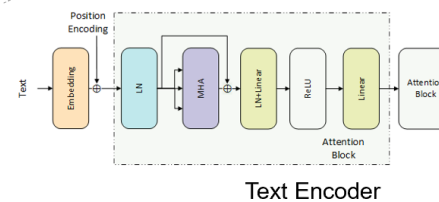
Output: Jet classification (10 classes)

Neural Network

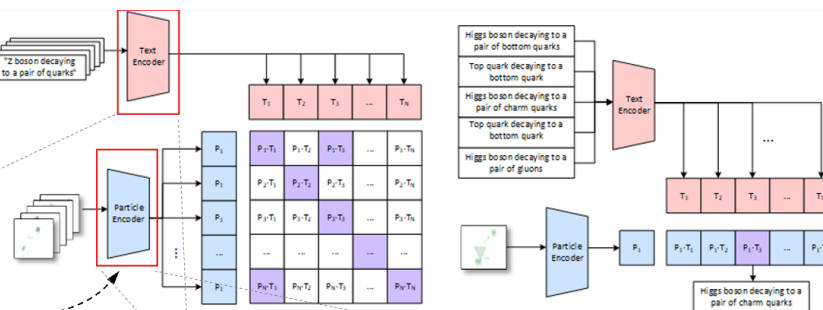
Original CLIP



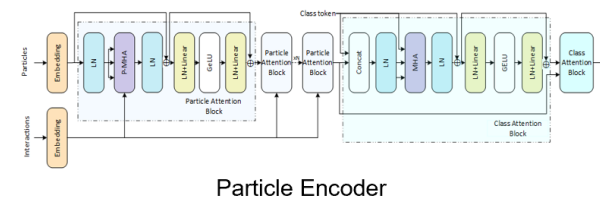
① Intro particle encoder



CLPP for HEP Data Pre-training

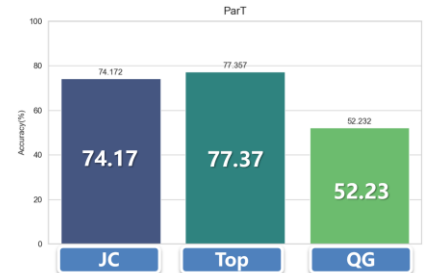


② Solving the problem of positive and negative sample errors when the text is not rich enough

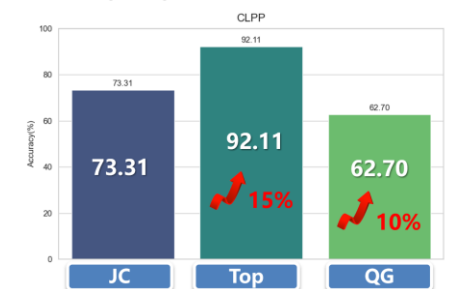


Result

Particle Transformer



CLPP (Ours)

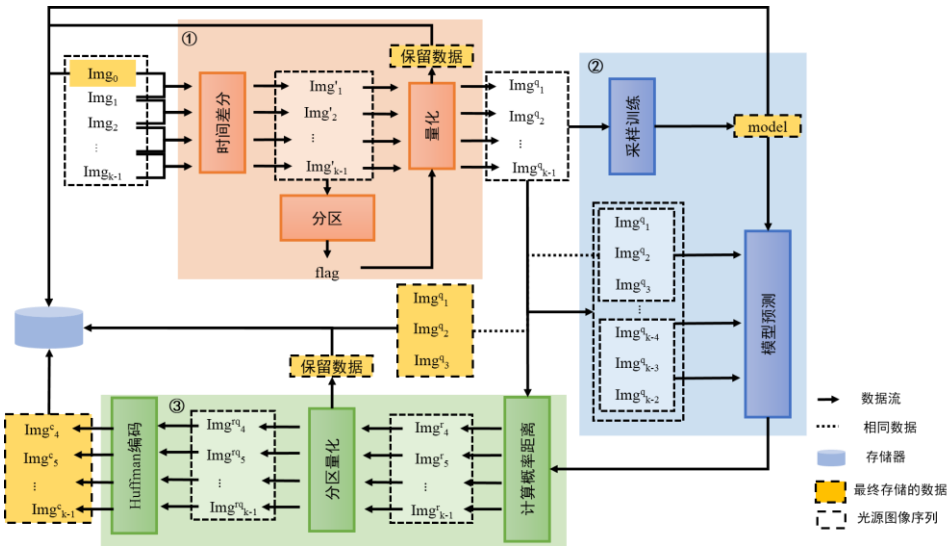


- **Generalization performance** improvement: contrastive learning improves cross modal prediction accuracy by 10-15% (compared to traditional supervised learning)
- Its **in-distribution performance** on the original test set slightly underperforms traditional supervised learning



- 大科学装置产生了海量的实验数据，对存储系统带来很大挑战
- 从理论上讲，熵率是给定数据源可能达到的最佳压缩比
- 传统无损数据压缩难以发现数据之间的长期依赖关系，因此难以实现更好的压缩比
- 斯坦福大学提出的DeepZIP压缩算法在基因数据等方面体现了更好的压缩率，但是压缩时间长
- 计算中心**面向光源图像数据**，采用差分、量化、AI训练、异构计算等方法，取得很好的效果
- 下一步规划
 - AI压缩与数据特点密切相关，将与大科学装置合作，训练不同的数据类型
 - 将该方法实际部署到计算平台中，对用户透明

	原始AI算法	优化的AI算法	JPEG2000	PNG	ZSTD
压缩率	2.08	1.78	1.26	1.43	1.13
时间(秒)	1281.4	3.7	0.8	0.6	0.6



- 基于平台深入研究机器学习和大模型技术的应用以提升网络性能分析、网络异常检测、网络威胁发现精准度

Step 1 : 网络和安全数据集构建

- 网络数据集
- 恶意流量数据集
- 钓鱼攻击数据集
- 网络安全运维数据集
- 高能物理软件漏洞数据集
- 网络安全标准数据集

Step 2 : AI方法应用

- 基于经典**机器学习**的网络威胁检测
人工特征工程，提取网络性能、攻击的统计属性
- 基于**深度学习**的网络威胁检测
自动抽取特征
- 基于**Transformer**的网络安全模型
预训练和微调

Step 3 : 网络和安全分析场景

- 网络异常检测
- 网络未知威胁检测
- 钓鱼攻击检测
- 网络与安全运维智能体构建
- 高能物理软件漏洞检测
- 安全框架构建和应用评估

基于规则的高危害已知攻击检测

基于行为特征的高级威胁检测

基于深度学习的恶意流量检测

基于大模型的技术体系研究

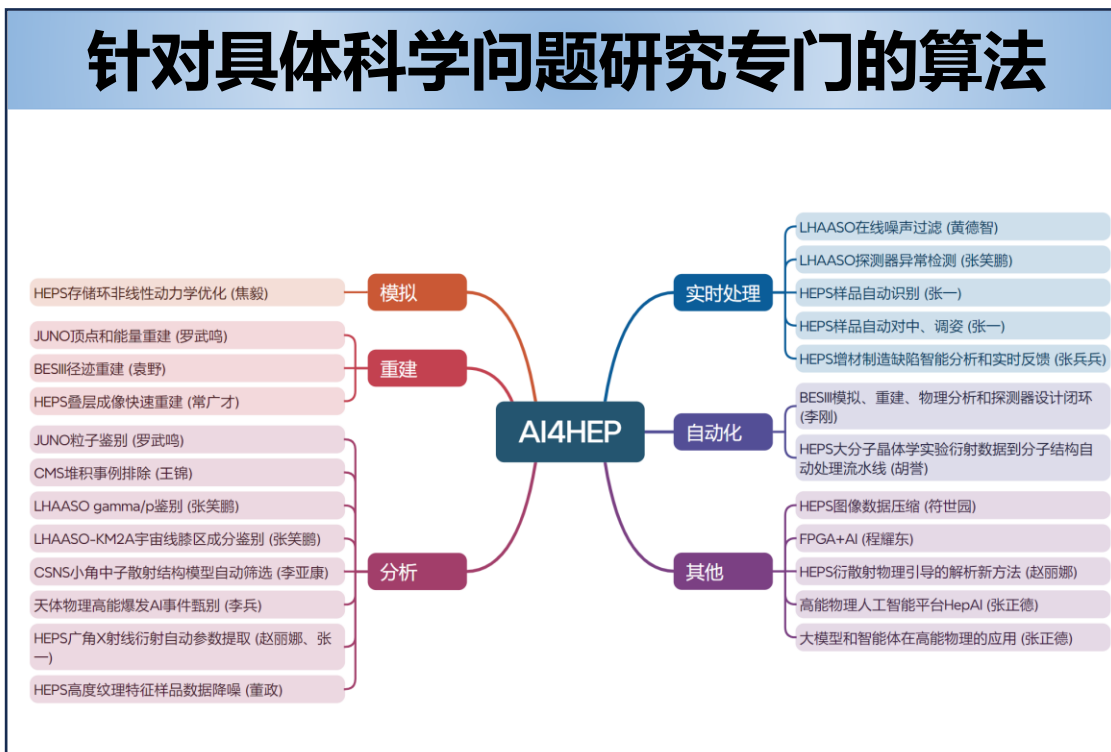
数据安全技术研究

基于小模型 → 基于大模型

- ✓ 深入分析大装置网络和安全场景特点，利用AI技术构建可靠的技术能力底座，实现**精准检测**和**高效响应**处置
- ✓ **降低人工和专业依赖**，提高大装置网络和安全智能分析和检测能力，为**大装置网络和安全能力提升**提供参考



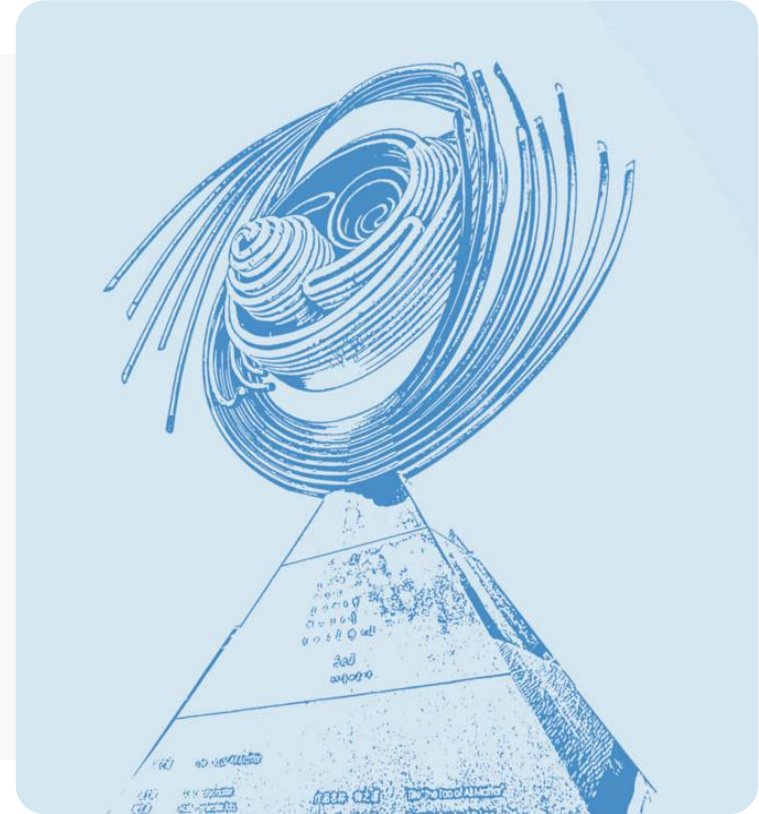
针对具体科学问题研究专门的算法



- ◆ 执行分类、回归、降噪、识别、异常检测等任务。
- ◆ 用于模拟、重建、分析、自动化等具体场景。

Part 3

AI Agents



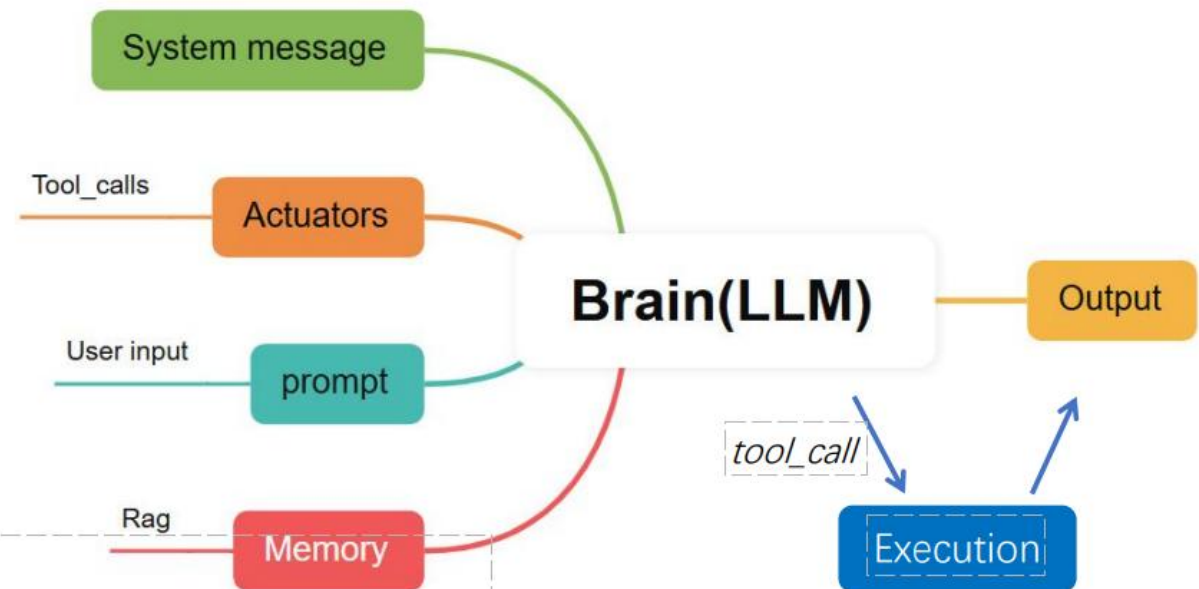


What is AI Agents?

An entity that can perceive its environment, make decisions, and take actions in order to achieve certain goals or sets of goals.

- Perception: multimodal -> text
- Decision: process text
- Action: reply or tool_call

- Agent = LLM + tools
- DIY reply_function to design the LLM output procedure



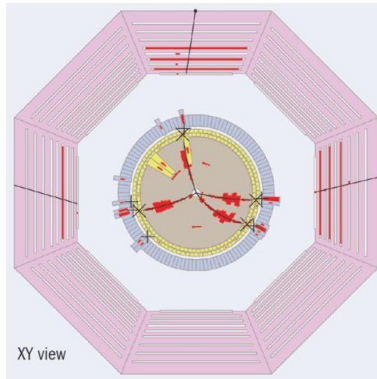
Physical analysis: a key link in scientific discovery



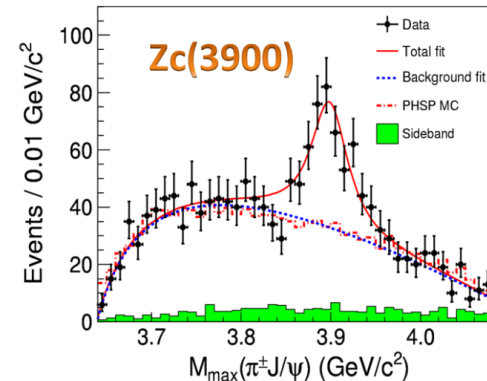
Data Processing Flow for High Energy Physics Experiments (Collider as an Example)



Collision data



Physical results



For the first time, mankind has discovered a particle composed of four quarks.

Complex tasks!

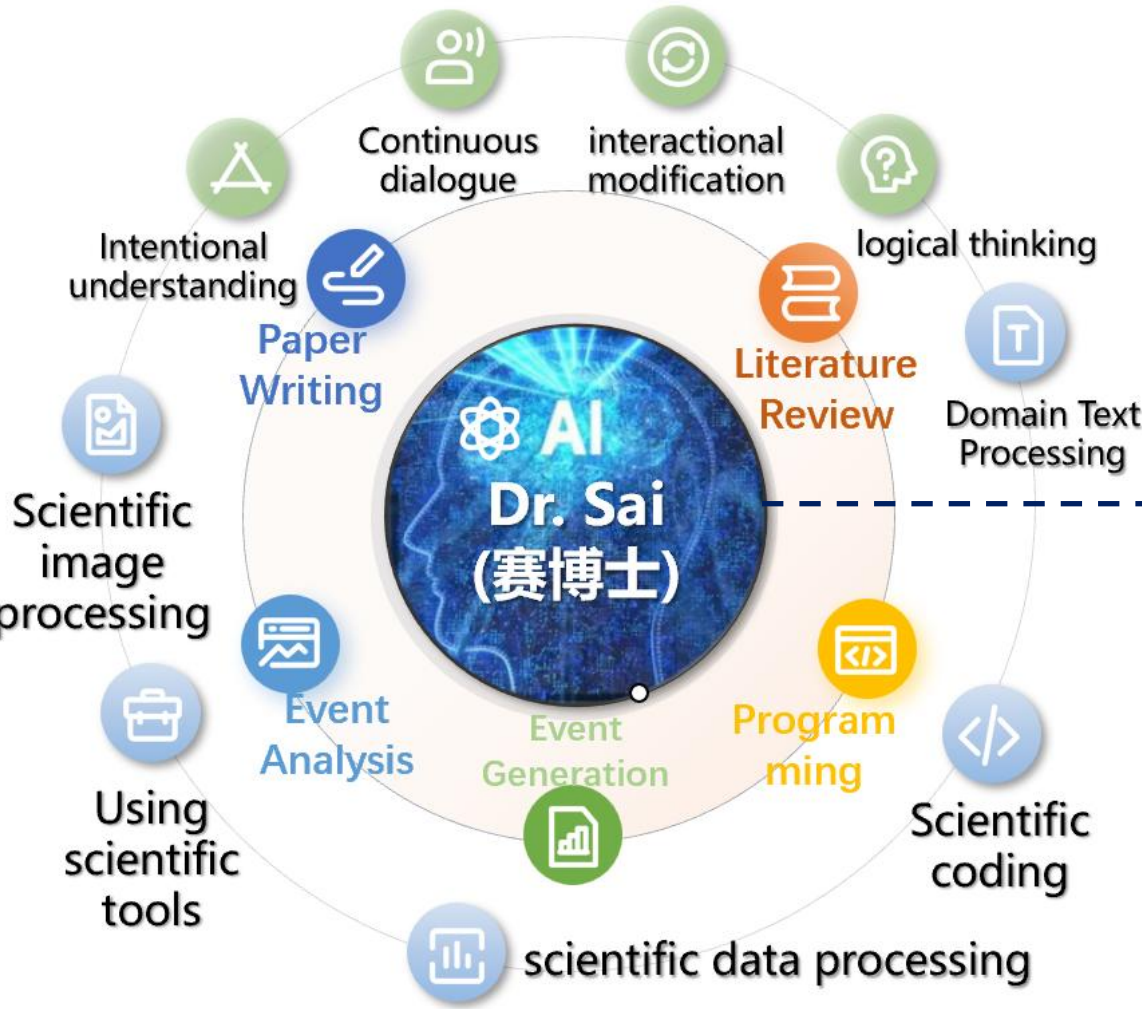
- Math + Physics domain expertise required
- Equivalent lines of code for windows/macOS
- A small mission requires: several people + several years

Dr.Sai Agents for Particle Physics Analysis

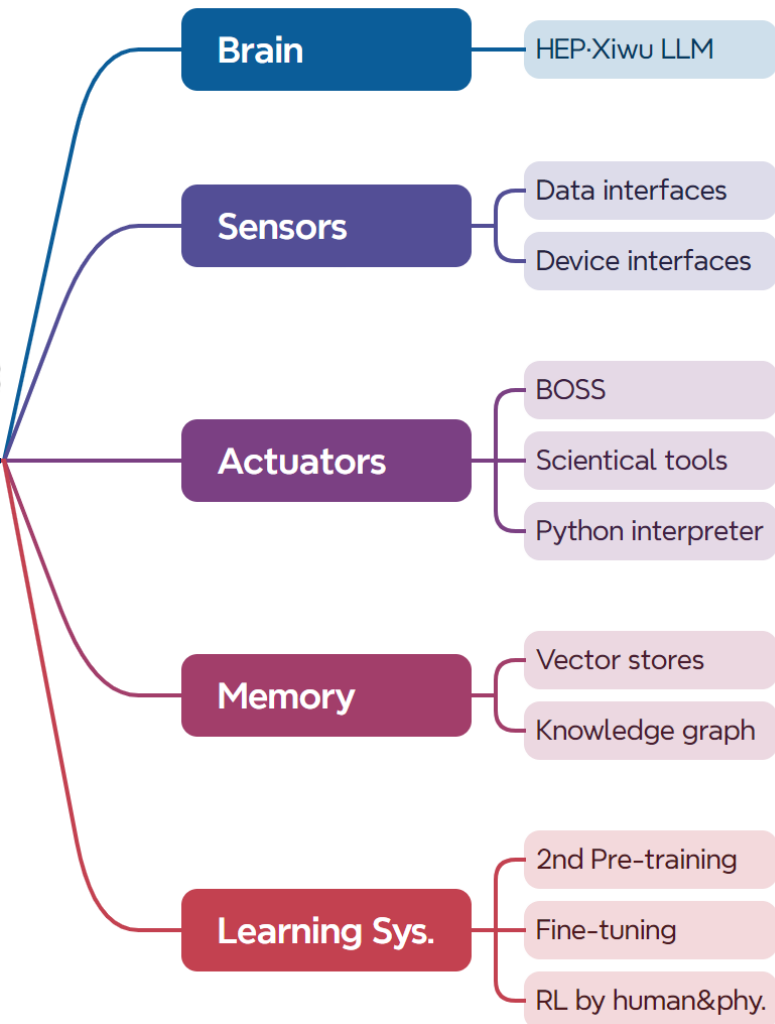


Let the large model conduct particle physics research

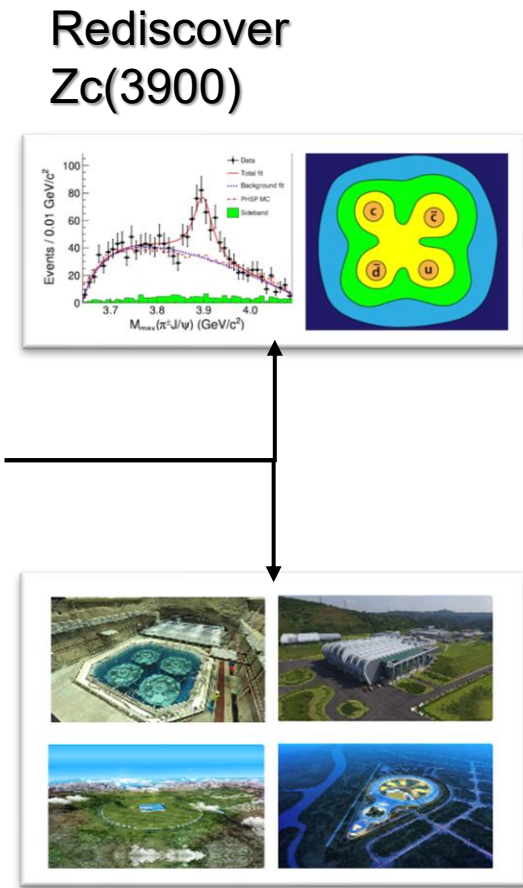
Essential: Modeling the research process.



Tasks & Required capabilities



Components



Applications 26

Latest Progress (July, 2025)



Open WebUI

DrSai

drsai.ihep.ac.cn

DrSai_BESIII_v2.0.0

有什么我能帮您的吗?

+

建议

J/psi -> mu+ mu-

psi(4260) -> K+ K- [J/psi -> e+ e-]

J/psi -> [rho+ -> pi+ pi0] pi-

$\psi(3770) \rightarrow \bar{p}p$

psi(4260) -> pi+ pi- [J/psi -> mu+ mu-]

psi(2S) -> K_S0 K+ pi-

INK

INK

INK

ink.ihep.ac.cn/xte...

(venv) [liaoy@aiboss001 17:36 ~/sharefs/DrSai/test]

\$ source start_worker.sh

gcc setup done

Use DrSai env: /hpcfs/bes/mlgpu/liaoyp/venv/bin/python3

[2025-07-10 17:36:28,799] [hepai/code-worker-v2-B0SS-8] [WARNING]: worker_cwd is not empty: /afs/ihep.ac.cn/users/l/liaoyp/sharefs/DrSai/test/run_test

[2025-07-10 17:36:28,799] [hepai/code-worker-v2-B0SS-8] [INFO]: Starting loop...

[2025-07-10 17:36:28,907] [worker_app.py] [INFO]: Worker register successfully: 'wk-6e2a91e0-4c6'

WorkerInfo(id='wk-6e2a91e0-4c6', type='common', network_info=WorkerNetworkInfo(host='0.0.0.0', port=42899, route_prefix='/apiv2', host_name='aiboss001.ihep.ac.cn', worker_address='http://202.122.33.201:42899/apiv2'), resource_info=[ModelResourceInfo(model_name='hepai/code-worker-v2-B0SS-8', model_type='common', model_version='1.0', model_description='This is a demo worker of HEP AI framework (HepAI)', model_author=None, model_owner=['liaoyp@ihep.ac.cn'], model_groups=['default'], model_users=['liaoyp@ihep.ac.cn'], model_functions=['__call__', 'get_dict', 'get_float', 'get_int', 'get_list', 'get_stream', 'hello_world', 'inspect_environment', 'inspect_system', 'interface', 'list_callable_functions', 'print_function_args', 'run_command', 'search_replace', 'write_code'])], status_info=WorkerStatusInfo(speed=1, queue_length=0, status='ready'), check_heartbeat=True, last_heartbeat=None, version='2.0', metadata={})

INFO: Started server process [3042774]

INFO: Waiting for application startup.

INFO: Application startup complete.

INFO: Uvicorn running on http://0.0.0.0:42899 (Press CTRL+C to quit)

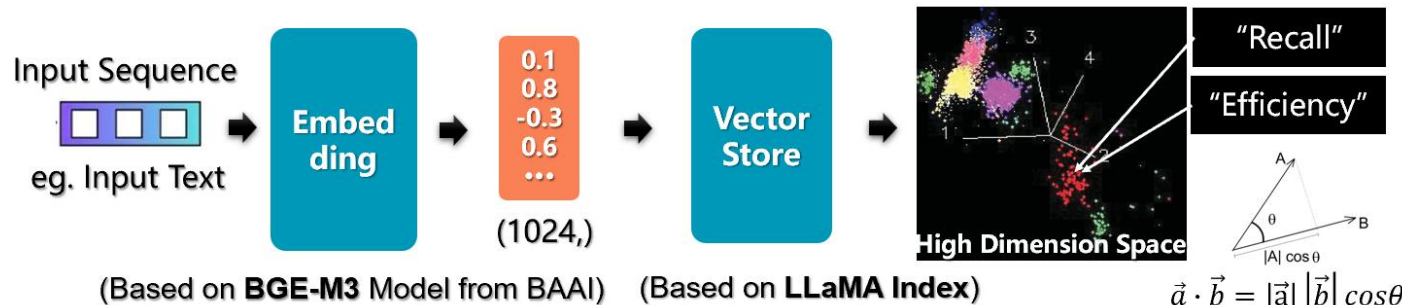
Online demo: <https://ihepbox.ihep.ac.cn/ihepbox/index.php/s/DsaHtndLiu52R5w>



Memory

Vector Databases and Knowledge Graphs

- ❑ Retrieval Augmented Generation (RAG)
 - ❑ The most cost-effective and promising solution for hallucination reduction
 - ❑ Store private data without retraining
- ❑ Current approach: vector storage
 - ❑ Embedding (using BAAI's BGE-M3 model):
 - ❑ Convert input data into high dimensional spatial vectors
 - ❑ RAG system (LLaMA Index).



What is HEPS?



HEPS stands for High Energy Photon Source. It is ...

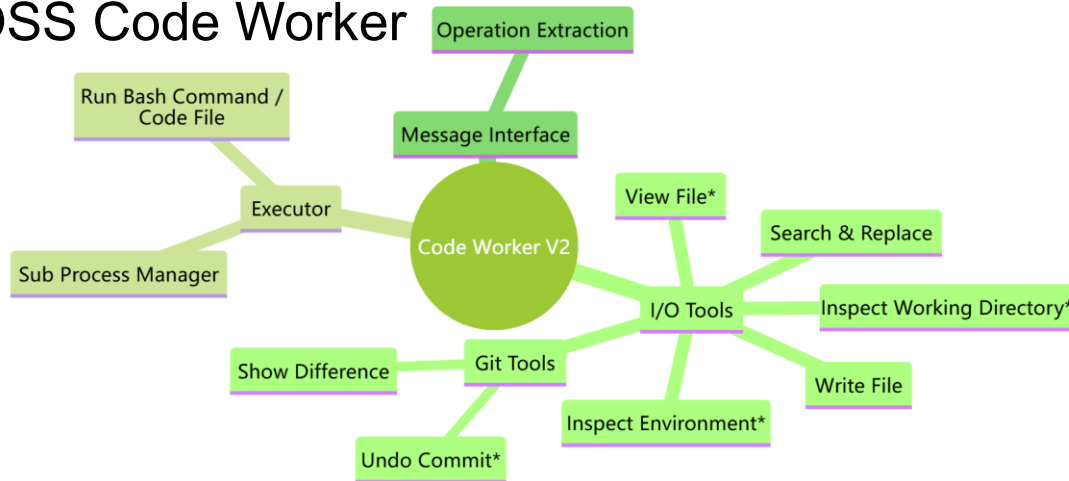
source: /docs/ihep/facilities.pdf

- ❑ Provide answers based on user intent and information
- ❑ Providing sources of information

Actuator

BOSS Code Worker

BOSS Code Worker



Run on HPC Cluster

- I/O Tools (files and data)
- Message interface
- Executor

HaiDDF



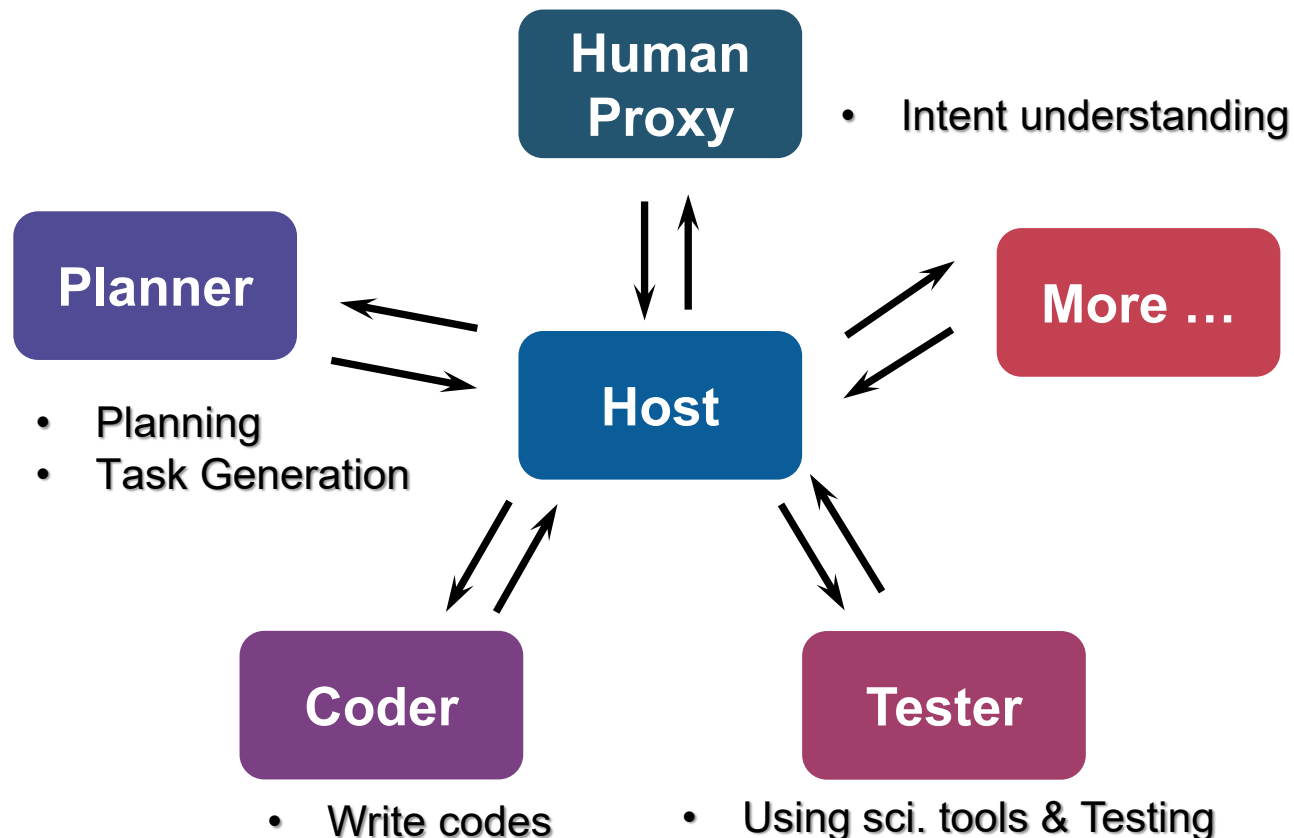
Dr.Sai

From **Q&A chatbot** to **expert assistant**



Multi agent dynamic task processing system supporting human-machine collaboration

- ❑ Multi-Agent Collaboration (**MAC**) is implemented based on [AutoGen](#) framework.
- ❑ Each agent is equipped with specific knowledge, tools, and LLM.
- ❑ A Host agent is introduced to manage group chats, making it easy to expand with more agents.
- ❑ A human proxy agent is introduced to allow humans intervene at any time.



- 🚀 **Technological breakthrough**
- 🔥 **Hierarchical task model:** The task tree system automatically decomposes complex targets and tracks the status and dependencies of subtasks in real-time.
- 🔥 **Domain customization process:** Optimize interaction logic for physical analysis scenarios.

Workflow:
Msg → host → LLM
Planner → **RAG** → LLM → host
Host → LLM



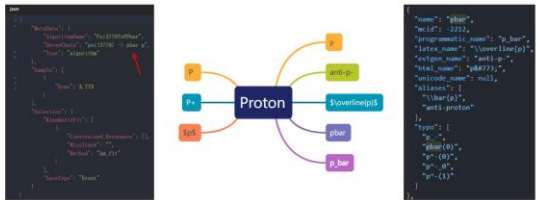
- **BOSS Mapping** – 解决BESIII领域代码生成不准确的问题
- **分层任务系统** – 适配BESIII物理分析任务多层次问题
- **守护进程** – 解决BESIII物理分析的长任务场景下的交互问题
- **知识图谱** – 实现基于关联关系的自我反思
- **拼写变体库** – 解决大模型时难以识别不同粒子拼写方式的问题
- **锚定词** – 通过提示工程提升智能体分配任务的准确率
- **消息卸载机制** – 解决多轮交互场景下智能体上下文爆炸的问题
- **运行时快照** – 智能体持续演进



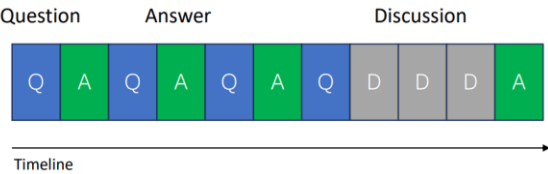
A database for compatible spelling variants and standardization of particle names

- Manual Data Annotation with **PDG** and **Particle** APIs

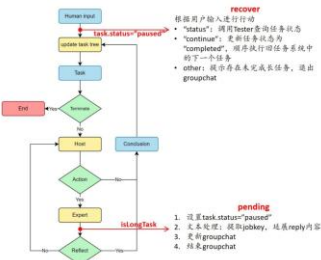
Usage: compatible with different spellings during IO, generate data for ML



- Host: sees **all** the QA pairs
- Experts: sees the **latest** QA pair and global info
- Global info: key info extracted from chat history



A typical conversation



Q: <task1> Expected messages
A: The jobs are submitted, please hold on.
Enter key words: ["status", "continue"] to proceed.
>> exit groupchat
Q: status
A: <job status from Tester>
>> exit groupchat
Q: continue
A: <answer to task2> (groupchat continue)
Q: <other query>
A: You have unfinished long tasks!
>> exit groupchat

BESIII分析算法生成

变量配置

分析算法模板

映射

可编译和执行的分析算法

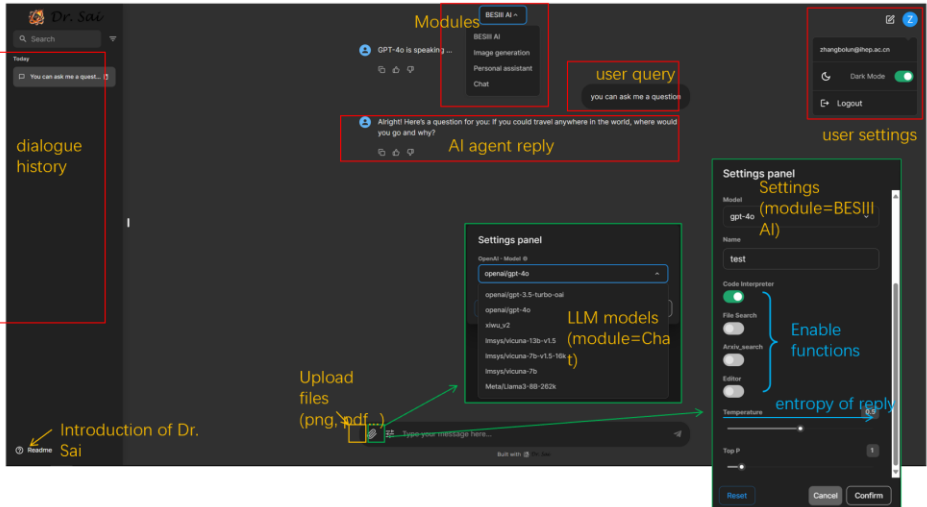
变量配置介绍

将所需标记减少至1/10~1/100
Pass@5: 43% → ~90%

Dr.Sai Agent - Interface



Dr.Sai v1.0 (July, 2024) (Based on Chainlit)



- Dr.Sai v2.x (Based on MagenticUI)
 - Features:
 - Data perception and progress details sidebar
 - Rich human-computer interaction
 - Expected to be launched on May 12, 2025
- Current version: <https://drsai.ihep.ac.cn>
- Dr.Sai v2.0 (2025.07) (Based on OpenwebUI)



- New features:
 - Automatic identity authentication
 - Guardian Process Long Task



- 我们期待高能物理人工智能科学家，他们目前已经实现了赛博士v2.0.0。
- 六大模块：大脑、感知层、执行层、记忆层、多智能体协作、人机交互界面
- 域适配技术：映射、代码工作者、DSL、拼写变体、锚定词、MOM、HTM、守护进程、Snashot
- 实现对**10个衰变通道的自动分析**，**未优化的新衰变通道的成功率约为30%**。



□ 数据融合智能体：AI-Ready数据的新构建方案





03 More Agents

更多智能体

01 Dr.Sai BESIII

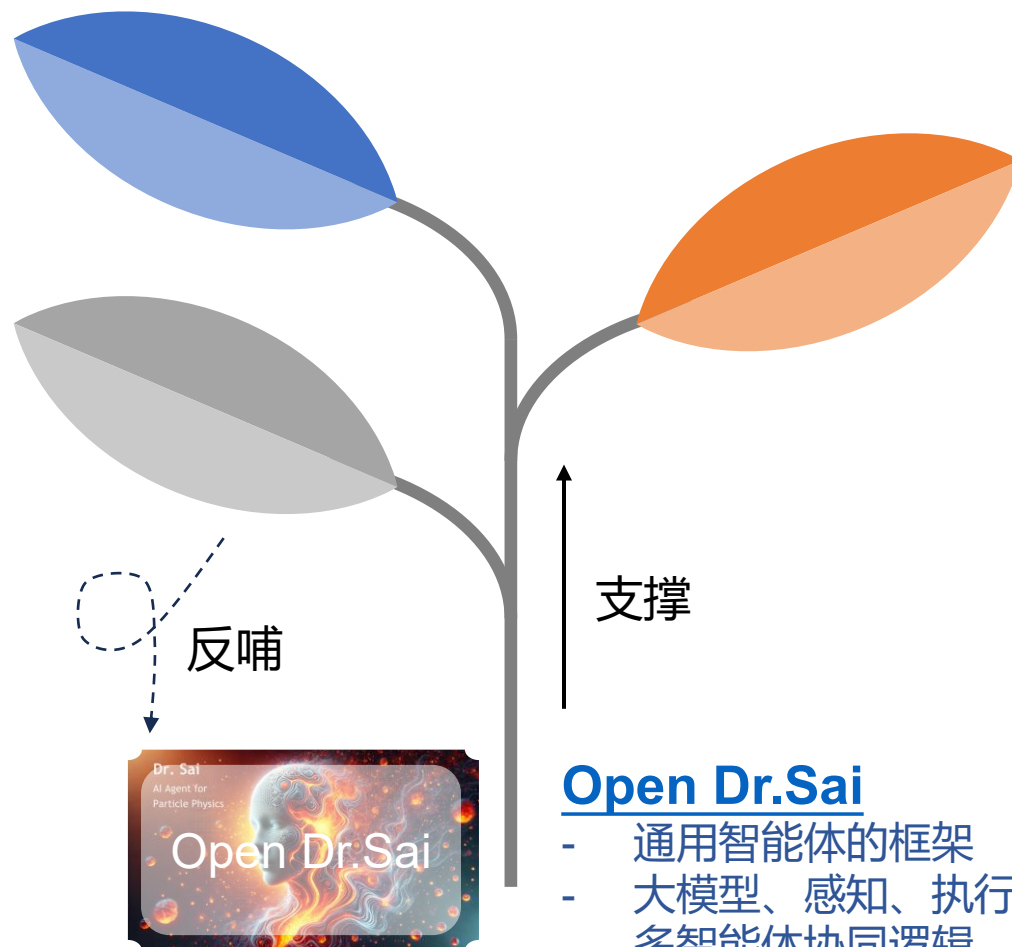
粒子物理分析智能体

- 增强物理能力
- 感知BESIII数据
- 搭载BOSS工具

02 HEPS Data Agent

光源数据智能体

- 增强数据能力
- 感知光源数据
- 搭载学科方法学工具



Open Dr.Sai

- 通用智能体的框架
- 大模型、感知、执行、记忆等基础组件
- 多智能体协同逻辑
- 权限验证及人机交互

<https://code.ihep.ac.cn/hepai/drsai>

《智能体构建实用指南》：

https://note.ihep.ac.cn/s/_vSnuf5Uh



◆ OpenDrSai 支持开发部署一体化 (Power by DDF2)

```
def create_agent() -> AssistantAgent:

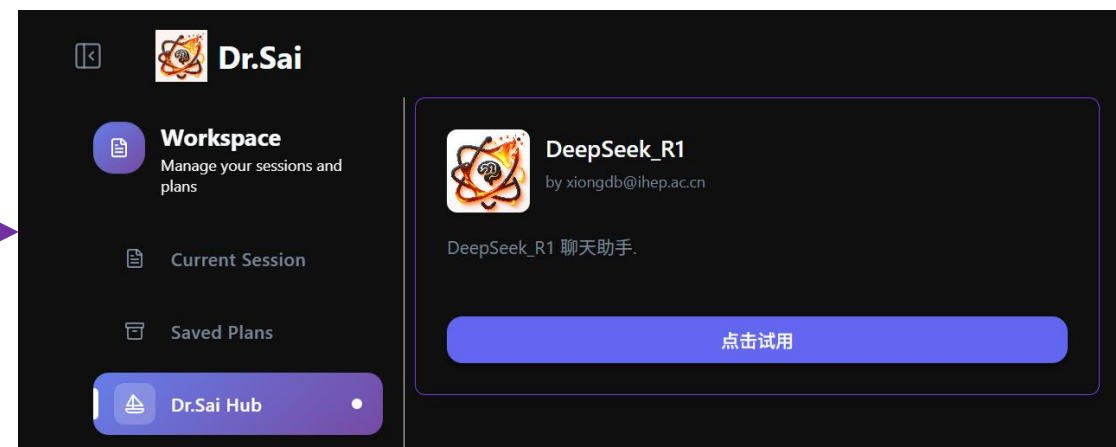
    # Define a model client. You can use other model client that implements
    # the 'ChatCompletionClient' interface.
    model_client = HepAIChatCompletionClient(
        model="deepseek-ai/deepseek-r1:671b",
        api_key=os.environ.get("HEPAI_API_KEY"),
        base_url="https://aiapi.ihep.ac.cn/apiv2",
    )

    # Define an AssistantAgent with the model, tool, system message, and reflection enabled.
    # The system message instructs the agent via natural language.
    return AssistantAgent(
        name="weather_agent",
        model_client=model_client,
        system_message="You are a helpful assistant.",
        reflect_on_tool_use=False,
        model_client_stream=True, # Enable streaming tokens from the model client.
    )
```

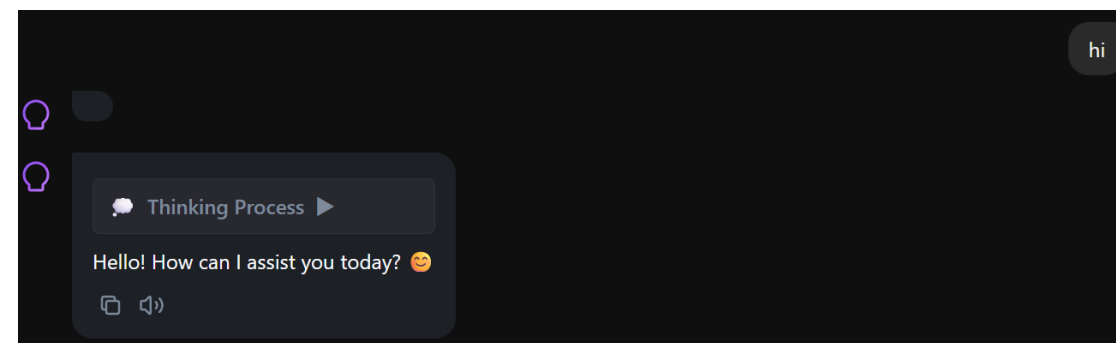
简单
定制

一键
部署

```
asyncio.run(
    run_worker(
        # 智能体注册信息
        agent_name="DeepSeek_R1",
        permission='groups: drsai; users: admin, xionddb@ihep.ac.cn',
        description = "DeepSeek_R1 聊天助手.",
        version = "0.1.0",
        logo="https://aiapi.ihep.ac.cn/apiv2/files/file-8572b27d093",
        # 智能体实体
        agent_factory=create_agent,
        # 后端服务配置
        port = 42812,
        no_register=False,
        enable_openwebui_pipeline=True,
        history_mode = "backend",
        # use_api_key_mode = "backend",
    )
)
```



开发及应用





■ 通过模型上下文协议MCP将科学工具搭载到智能体中

```
from drsai import AssistantAgent, HepAIChatCompletionClient, StdioServerParams, mcp_server_

tools = []

# Web fetch MCP tools
tools.extend(
    await mcp_server_tools(
        StdioServerParams(
            command=config["uvx"],
            args=config["run", "mcp-server-fetch"],
            env=None)))

# Self-def tools
@async def get_weather(city: str) -> str:
    """
    73 degrees and Sunny."

l agent who can talk to users about the weather and web
```

类型	描述	示例
数据工具	使智能体能够检索执行工作流所需的上下文和信息。	读取PDF文档，或进行网页搜索，查询数据库或CRM系统。
操作工具	使智能体能够与系统交互以执行操作。	调用科学工具、解析结果，向数据库添加新信息、更新记录或发送信息，发送电子邮件和短信，将客户服务工单转交给人工处理。
编排工具	智能体本身可以作为其他智能体的工具——参见“智能体编排”部分中的“管理者模式”	研究智能体、写作智能体、退款智能体。

The Role of Data Agent



Agent

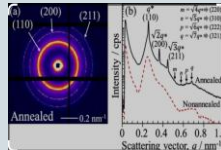
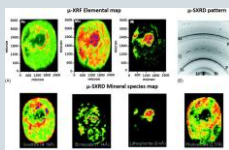
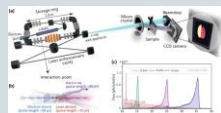
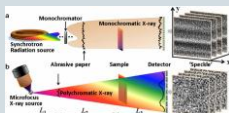
Utilizing AI intelligent agent technology to process and supply data from HEPS&CSNS.

Data Management System (Domas)

Experimental data

Metadata

Spectroscopy, Diffraction, Scattering, Imaging, etc.



True Label

① Update

- Proposal
- Experiment
- Materials
- etc.



Knowledge Base

Sensor

③ Data Recognition

- Format Conversion
- Execute commands

④ Automatic processing

- Execute tools or code

⑤ Verification and Storage



API

⑥ Data for multi-objective research

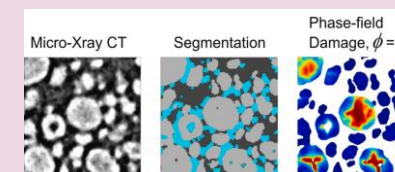
Task

② Task identification

- Reconstruction, Peak finding
- Denoising, Compression
- Segmentation, etc.

Actuator (Daisy)

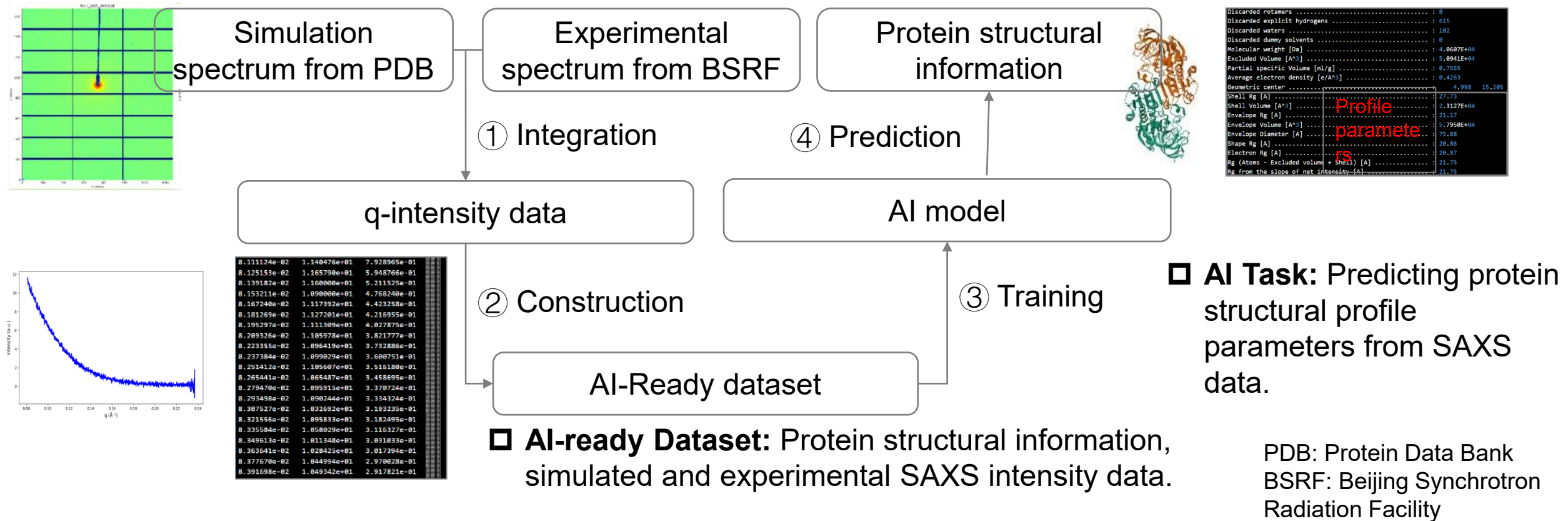
- XAFS tools
- XRD tools
- PDF tools



AI Agent for SAXS data



Experiments and simulations AI dataset construction for **Small-Angle X-ray Scattering (SAXS)** data



✓ A small-scale, end-to-end pipeline has been established.

➤ Currently, simulations are being expanded to generate a larger dataset for training higher-accuracy models.

AI Agent for Diffraction Data (Simulation)

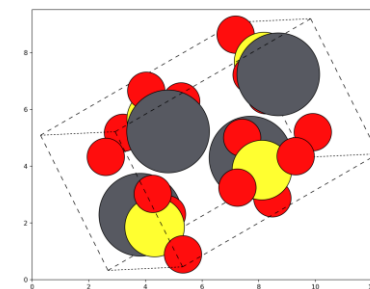


Building a Cross-Instrument X-ray Diffraction and Neutron Powder Diffraction Simulation AI Dataset.

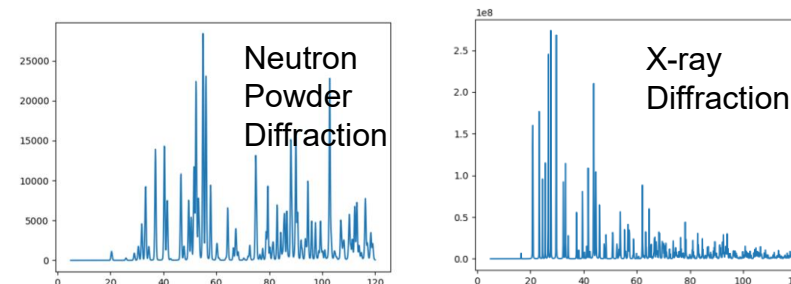
① Input: Crystal structure (CIF file):

请直接对当前目录下的PbSO4的CIF文件进行XRD模拟

Visualization:
Input structure



Visualization: Output diffraction spectrum



② Thinking and Simulation Process

SCI-DataAgent Today at 14:44
SCI_DataAgent发言:
思考用时 19 秒
用户需要对PbSO4的CIF文件进行XRD模拟, 需调用gsasii_xrd_process_agent的xrd_simulate_gsasii_post功能。已确认CIF文件路径为: /aifs/user/home/xionddb/VSproject/mutil_mcp_tools/files/thread_22578926-f5e3-48ef-873b-13a8fe7ca3e4/XRD_Neutron_power_diffraction/PbSO4-Wyckoff.cif。需要规划具体参数和执行流程。

由 planner_agent 进行接下来的操作, 现在由 planner_agent 接手。

planner_agent发言:

思考用时 40 秒

任务计划如下:

| --- | --- | --- |

| 任务专家 | 任务内容 | 前提条件 |

| gsasii_xrd_process_agent | 调用GSAS-II的XRD模拟功能对PbSO4-Wyckoff.cif进行XRD图谱模拟 | |

gsasii_xrd_process_agent发言:

X-Ray Diffraction Simulation Results

| --- | --- |

| Item | Value |

| status | success |

| project_file | /aifs/user/home/xionddb/VSproject/mutil_mcp_tools/files/thread_22578926-f5e3-48ef-873b-13a8fe7ca3e4/XRD_Neutron_power_diffraction/PbSO4_XRD_inst_d1a/PbSO4_XRD.gpx |

| data_file | /aifs/user/home/xionddb/VSproject/mutil_mcp_tools/files/thread_22578926-f5e3-48ef-873b-13a8fe7ca3e4/XRD_Neutron_power_diffraction/PbSO4_XRD_inst_d1a/PbSO4_XRD_data.csv |

| reflections_file | /aifs/user/home/xionddb/VSproject/mutil_mcp_tools/files/thread_22578926-f5e3-48ef-873b-13a8fe7ca3e4/XRD_Neutron_power_diffraction/PbSO4_XRD_inst_d1a/PbSO4_XRD_refl.csv |

| plot_file | /aifs/user/home/xionddb/VSproject/mutil_mcp_tools/files/thread_22578926-f5e3-48ef-873b-13a8fe7ca3e4/XRD_Neutron_power_diffraction/PbSO4_XRD_inst_d1a/PbSO4_XRD.png |

| error | None |

③ Output: Simulated diffraction pattern

❑ **AI-ready Dataset:** Crystal structures, simulated X-ray and neutron powder diffraction data.

✓ Simulations for selected crystal structures have been completed.

➤ Currently collecting lithium-ion battery material-relevant structures to construct the AI-ready dataset for battery-focused AI tasks.

Agent for 信息化 (内部)

孙智慧



- 采用分层架构设计，包括感知层、认知+决策层、执行层、记忆层、AI服务应用
- 增加信息化服务智能体到规划的信息化架构中，部署开源大模型（**DeepSeek-R1**、**Qwen2.5**）完成认知+决策层的实现，设计基于语音+图像识别的全新前端简化与用户的交互，通过NLP、知识图谱、强化学习服务于上层AI应用



小结：ChatGPT应用于科学研究的思考（2023.03）



直接应用

1. 写代码，代码纠错，通用知识问答。
2. 写综述、写摘要、翻译、词句润色、语法纠正…（注：已引起争议讨论）

潜在应用

1. **领域专业微调模型**，应用于大科学装置等，辅助装置操作，提升智能化水平。例如：在实验前辅助设计实验、在实验中辅助运行控制、在实现后辅助数据分析。
2. **利用大模型无监督预训练方法让AI把握所有数据中的全局规律**，例如：在探测类装置中，使用预训练大模型的方式进行自监督学习，以少量的异常样本为Prompt对模型进行提示，使模型能对原始数据进行质量判别，自动发现异常数据(感兴趣数据)。
3. **引导ChatGPT使用科学工具**：计算器、积分工具、科学领域分析工具等。
4. **科学领域专业知识检索**，存储在LLM中的知识可以显著提高知识密集型任务的性能。
5. **分布外(Out-of-Distribution, OOD)泛化**，传统的微调可能会过拟合训练集并且有较差的分布外泛化能力；而少样本的上下文学习 (in-context learning) 能够有更好的分布外泛化性。
6. **大模型的Prompt来做物理知识嵌入**。

科学上要求精确、可解释、误差性依然还是挑战，知识更新到模型更新的延迟带来的错误

Part 4

Common Infrastructure



- A hub hosts data, algorithms, computing power, and applications

高能人工智能平台HepAI



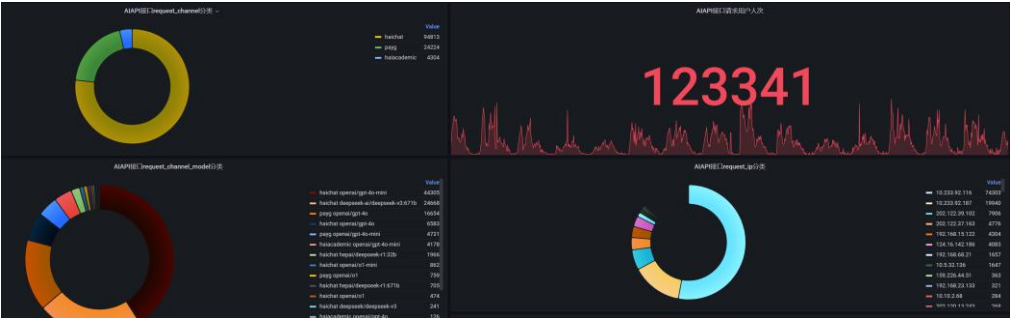
多学科场景的AI基础设施与协同引擎

核心定位

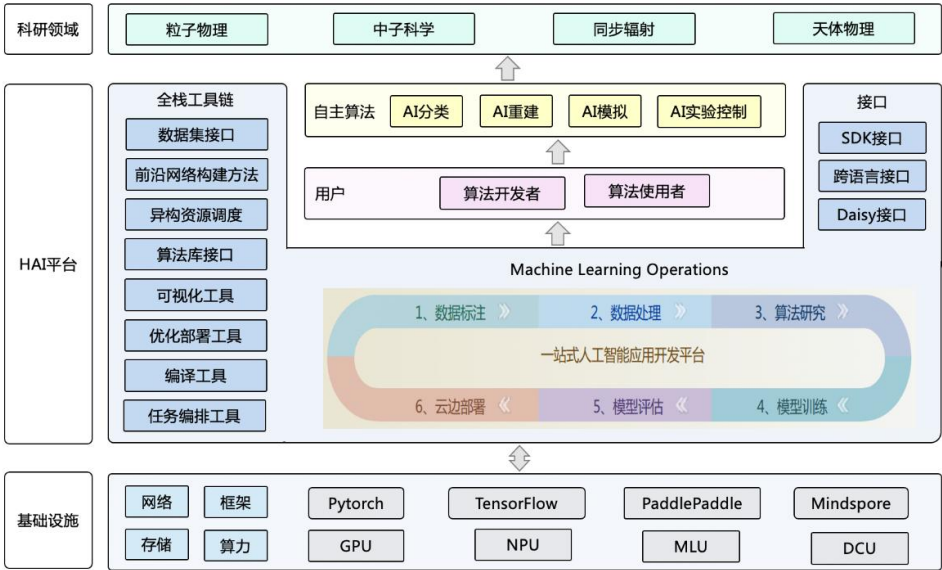
- ✓ **共性基础设施**：面向粒子物理、粒子天体物理、同步辐射、中子科学、加速器等多领域，提供从数据治理、算法迭代到算力调度的一体化AI研发闭环。
- ✓ **软件定义平台**：以AI模型流为核心，打通“数据-算法-算力-协作”全链路，降低跨学科AI应用门槛。

平台价值

- **承载AI模型/工具**：已支持领域50+模型/工具。
- **新建专用AI算力 (2024)**：5PFlops (16卡GPU+ 48卡DCU)，200TB全闪存硬件
- **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。

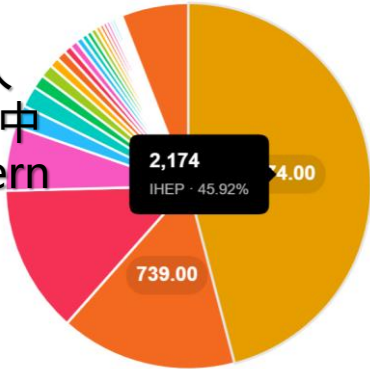


<https://ai.ihep.ac.cn>



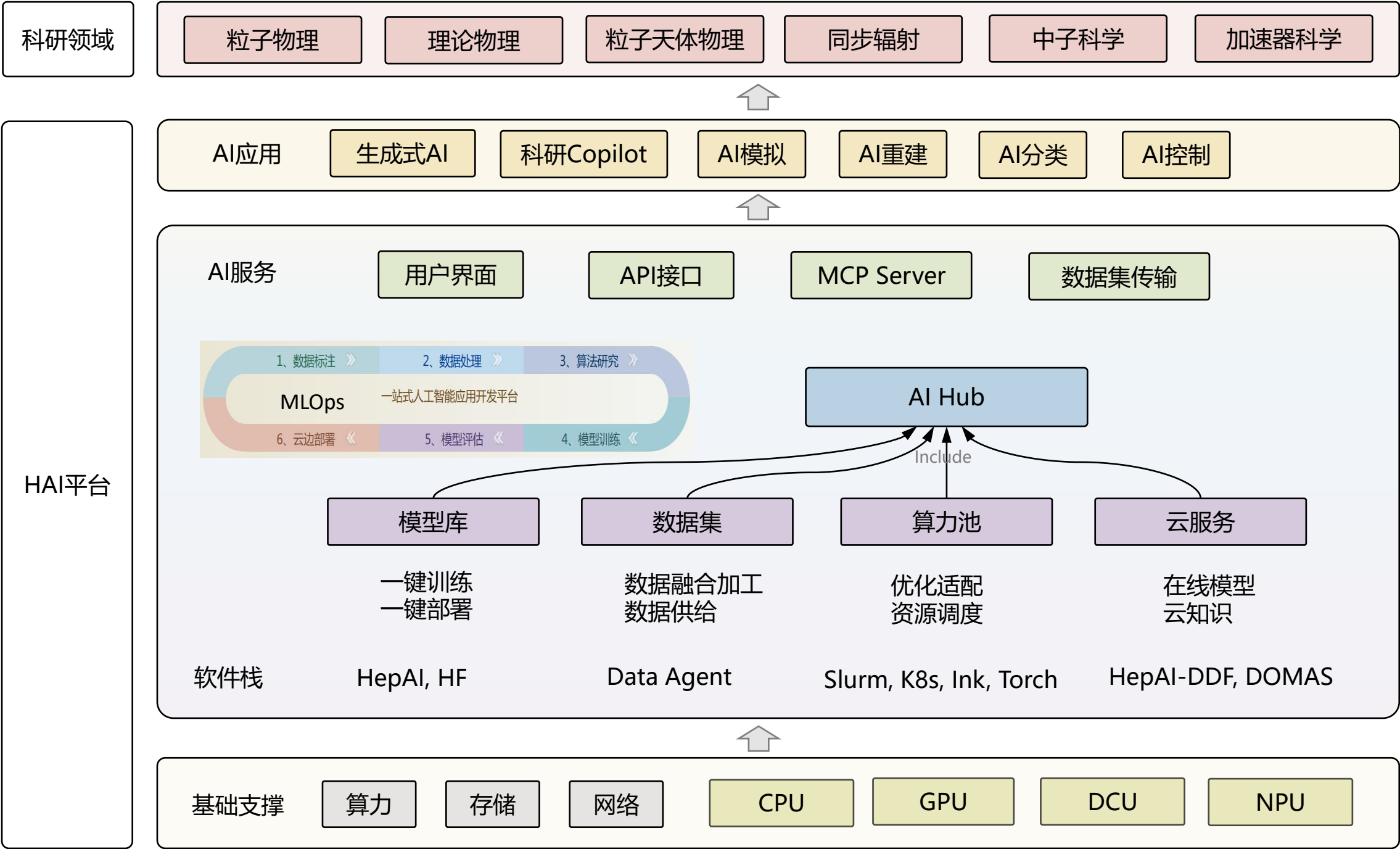
HepAI平台架构

- **用户总数：8000 人**
 - ◆ 高能所：约3000人
 - ◆ 中科大、国科大、中山大学，北大，Cern
- **日均活跃：800人**



过去15天，模型请求人次10万+

架构





高能AI框架

HepAI

高性能AI分布式框架Hai-DDF

`pip install hepai`

🔧 核心功能

✓ 统一访问：跨异构资源（CPU/GPU/DCU）无缝集成AI模型、科学工具、矢量数据库和知识图谱。

✓ 智能调度：自动心跳监测+动态负载平衡。

✓ 统一的API网关：具有API-KEY身份验证的标准化服务接口，可增强安全性。

🔥 远程模型调用：

NEW 对远程服务器模型部署的突破性支持，提供无缝的本地呼叫体验。

🔥 适配MCP：模型上下文协议，访问多个+MCP服务器。



- 50+模型
- 2个本地工具
- 多种MCP工具
- In one server

□ HaiDDF2.0 was lanched in Feb. 2025.



■ 支持55+在线模型/工具

- deepseek-v3, r1
- GPT系列 (4o, 5, o1, o3) 等
- 千问系列
- Claude系列
- 图像模型
- 本地deepseek
- 本地boss code worker等

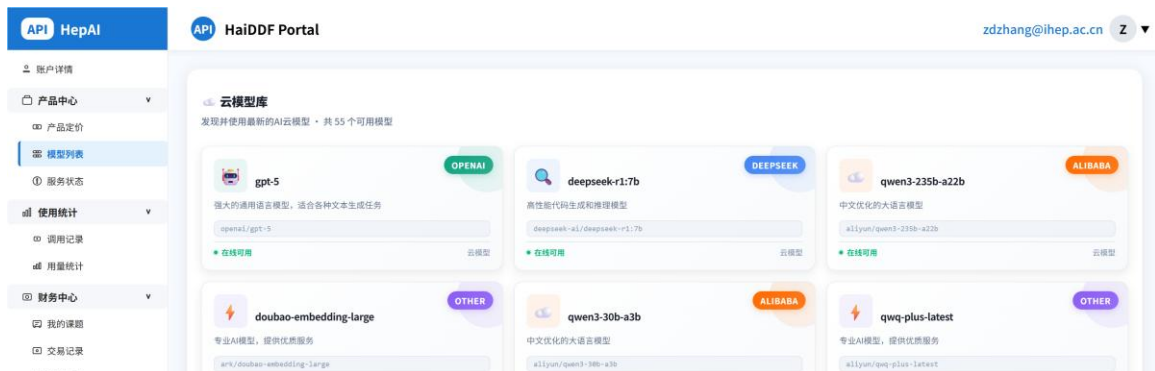
■ 通过API发起网络请求调用

```
from hepai import HepAI
```

```
client = HepAI(  
    base_url="https://aiapi.ihep.ac.cn/apiv2",  
    api_key=os.getenv("HEPAI_API_KEY")  
)
```

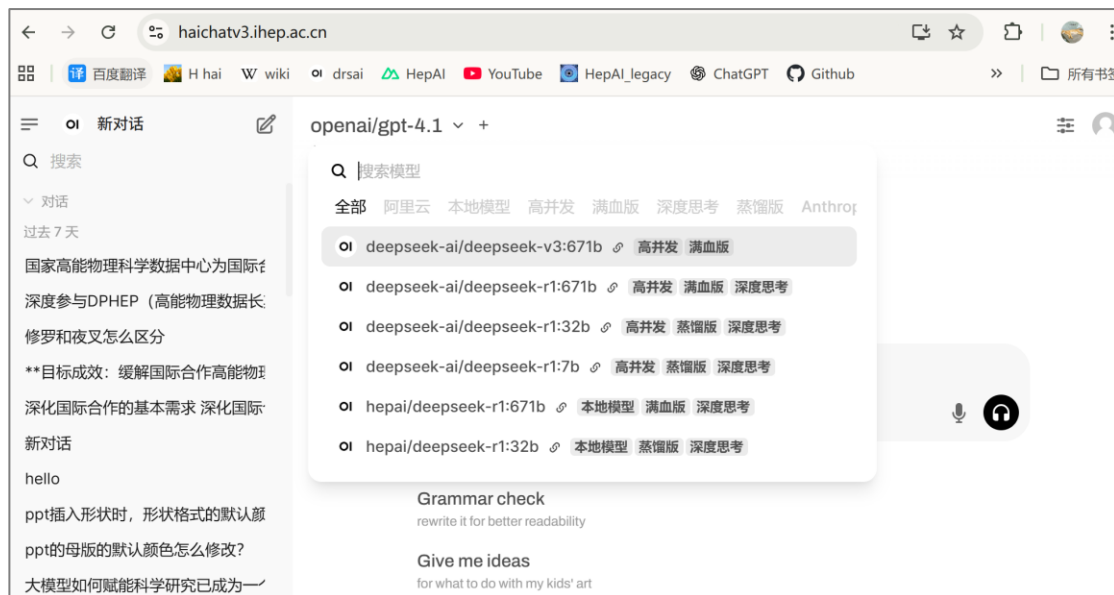
```
response = client.chat.completions.create(  
    model="hepai/deepseek-r1:671b",  
    messages=[  
        {"role": "user", "content": "你好"}  
    ],  
    stream=True  
)
```

```
for chunk in response:  
    if chunk.choices[0].delta.content:  
        print(chunk.choices[0].delta.content, end="", flush=True)
```



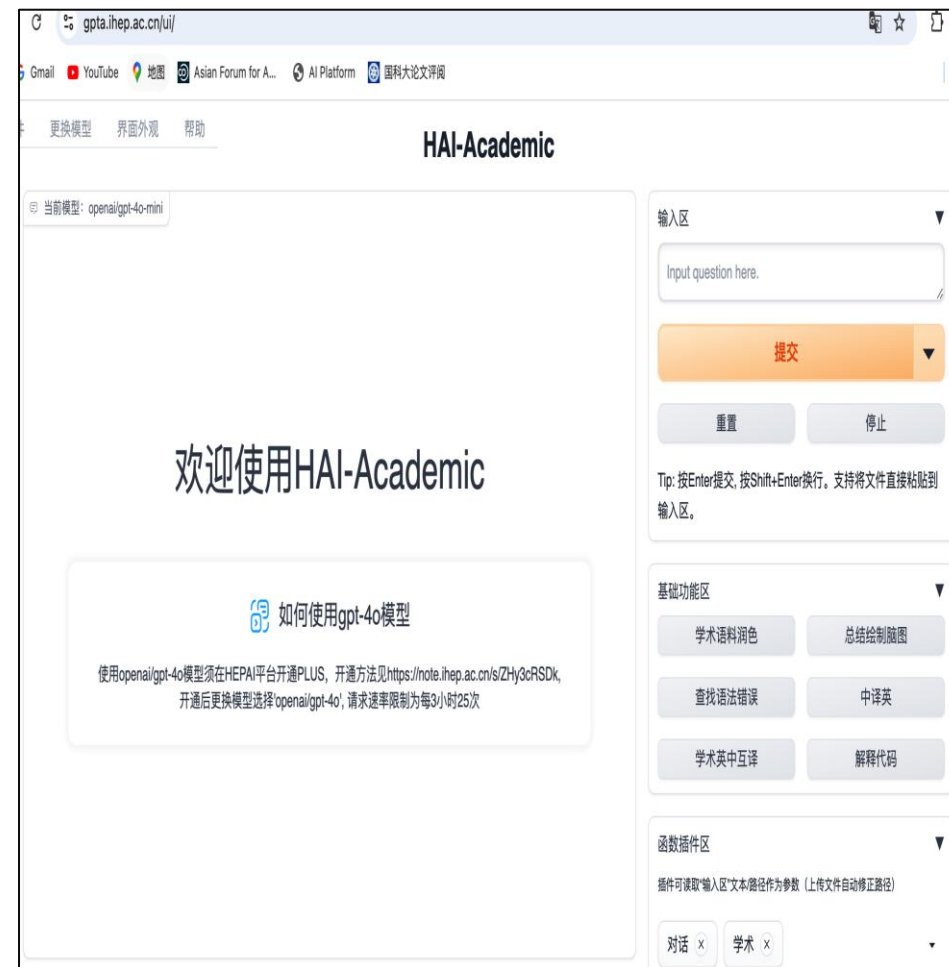


■ 生成式人工智能服务HaiChatv3（张易于）



<https://haichatv3.ihep.ac.cn>

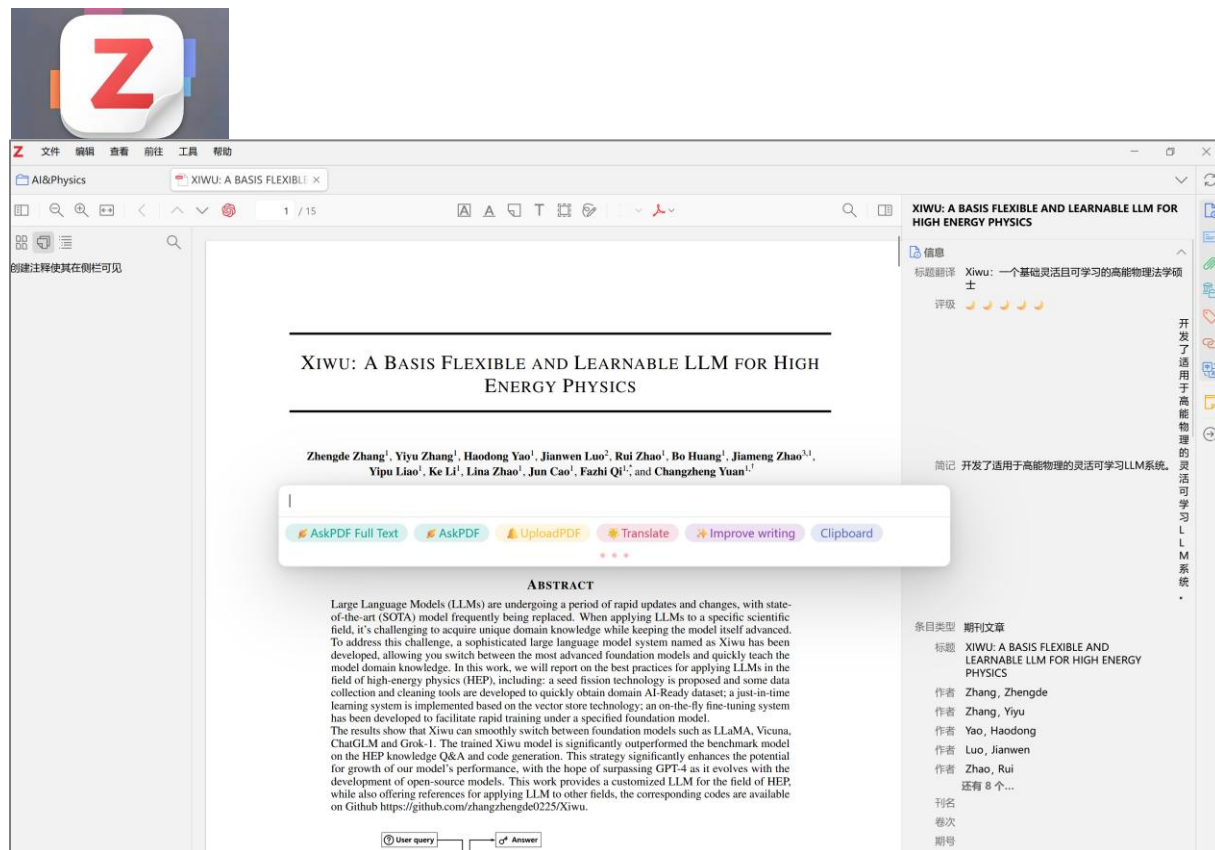
■ 学术论文助手HaiAcademic（姜法宇）



<https://gpta.ihep.ac.cn>（内部）



■ 文献管理和AI辅助阅读工具



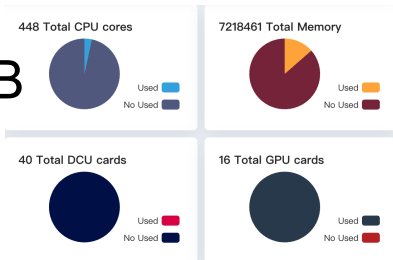
■ AI辅助编程：Claude Code



■ 通过插件和配置HepAI API-KEY实现



- 共计80张卡：8张NV L40，8张NV A800，48张DCU K100ai，16张NPU 910B
- 其他非AI专用~300张卡，以V100为主
- 潜在算力：40+NV 5090，100+SUGON DCU，100+Huawei NPU



用途	设备名称	数量（台）	配置
计算	A800 GPU服务器	1台	8 * A800 80GB PCI-e NVIDIA GPU卡 2 * Intel(R) Xeon(R) Gold 6430(32 core) 1TB 内存，7.68 TB NVME 本地硬盘 2 * 200Gbps IB网卡，1 * 25Gbps 以太网卡
	L40 GPU服务器	1台	8 * L40 48GB PCI-e NVIDIA GPU卡 2 * Intel(R) Xeon(R) Gold 6430(32 core) 1TB 内存，7.68 TB NVME 本地硬盘 2 * 200Gbps IB网卡，1 * 25Gbps 以太网卡
	DCU服务器	6台	8 * K100AI 64GB PCI-e 国产海光DCU卡 2 * Intel(R) Xeon(R) Gold 6430(32 core) 1TB 内存，7.68 TB NVME 本地硬盘 2 * 200Gbps IB网卡，1 * 25Gbps 以太网卡

■全闪，可用容量200TB

■个人目录

事项	说明
路径	/aifs/user/home/<username>
用途	存放个人datasets/models等文件
权限	用户个人可读/写
限额(quota)	默认值500GB、30万个文件，可按需修改

■公共目录

事项	说明
路径	/aifs/public/data
用途	存放公开datasets/models等文件
权限	普通用户可读、不可写，公共数据管理员可读/写
限额(quota)	按需定制



- 使用ECS弹性服务器
 - 弹性云服务器（Elastic Cloud Server, ECS）为用户提供可直连、可扩展的算力服务器
 - 允许用户动态调整计算资源的配置，包括CPU、内存、GPU卡等，满足灵活的AI算力需求
 - 主要用于单机单卡、单机多卡的AI模型调试、训练和部署
 - 允许用户像使用个人服务器一样直接与算力资源进行交互
 - 优势，即支持对AI模型和软件代码的**运行时调试**

① 输入`hai-ecs`申请计算资源

```
(base) [zhangyiyu@ailogin001 ~]$ hai-ecs
[DEBUG] Response text: {"status":200,"msg":"请求成功","data":[]}
Applying Elastic Cloud Server (ECS) job with the following configuration:
  num_nodes: 1
  cpu cores: 8
  memory: 64G
  accerlerator_cards: gpu * 1

[DEBUG] Response text: {"status":200,"msg":"作业提交成功","data":{"jobId":2437,"jobType":"enode","jobPath":"/aifs/user/home/zhangyiyu/.ink/Jobs/enode-20250717-222152"}}
Job submitted, job_id: `2437`
[DEBUG] Response text: {"status":200,"msg":"请求成功","data":[]}
Waiting for the job to be running ... 0.06s.[DEBUG] Response text: {"status":200,"msg":"请求成功","data":{"clusterId":"slurm","jobId":2437,"jobStartTime":"2025-07-17T22:21:53","jobStatus":"RUNNING","jobNodeList":["aigpu001","jobSubmitTime":"2025-07-17T22:21:53","jobType":"enode","connect_sign":"True"]}}
Job (id=`2437`) current status: RUNNING

The ECS is ready!
HostName: ai.ihep.ac.cn
User: zhangyiyu
Port: 53307

For more information, please visit: `https://ai.ihep.ac.cn/docs`
You can connect to it via: `ssh -o UserKnownHostsFile=/dev/null zhangyiyu@ai.ihep.ac.cn -p 53307`
```

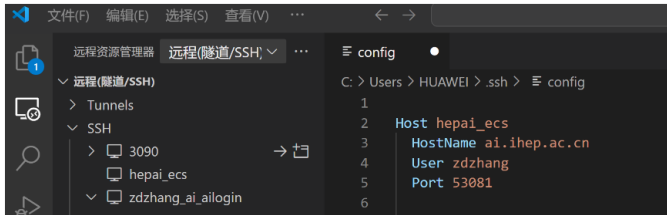
② 本地VSCode直连

本地VSCode直连ECS

在本地VSCode安装 Remote - SSH 插件后，可以实现直连ECS。

点击 远程资源管理器 - SSH - 打开SSH配置文件，编辑 ~/.ssh/config 文件，添加ECS的连接信息。配置如下：

```
Host hepai_ecs
  HostName <ECS_IP>
  User <username>
  Port <ECS_PORT>
```



配置项说明： - hepai_ecs 为自定义远程服务器名，可更改 - HostName 为ECS的IP地址，为 ai.ihep.ac.cn - User 为AI集群用户名 - Port 为ECS的端口号，此处不是默认 22，需要从 hai-ecs 命令获取，不同用户端口号不同，每位用户端口号固定。

连接ECS： - 配置完成后，在 远程资源管理器 的 SSH 下点击刷新，找到 hepai_ecs，点击连接，输入密码，连接成功。

注： - ECS为内部服务器，需在内网环境下连接，在外网环境下需提前打开VPN：vpn.ihep.ac.cn。



- 代码逐行解读
- 代码快速复现
- 代码托管
 - Gitlab
- 讲座、报告
 - ML研讨会
- 公开课程



6 Zheng-De Zhang

7668 views



2 年前

6082 views

- HepAI平台入门实践教程

如何通过网页UI与大语言模型聊天

Multi-Head Attention Module

This computes scaled multi-headed attention for given `query`, `key` and `value` vectors.

$$Attention(Q, K, V) = softmax_{seq} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

In simple terms, it finds keys that matches the query, and gets the values of those keys.

It uses dot-product of query and key as the indicator of how matching they are. Before taking the *softmax* the dot-products are scaled by $\frac{1}{\sqrt{d_k}}$. This is done to avoid large dot-product values causing softmax to give very small gradients when d_k is large.

Softmax is calculated along the axis of of the sequence (or time).

- `heads` is the number of heads.
- `d_model` is the number of features in the query, key and value vectors.

Number of features per head

Number of heads

These transform the `query`, `key` and `value` vectors for multi-headed attention.

```
69 class MultiHeadAttention(nn.Module):  
  
90     def __init__(self, heads: int, d_model: int, dropout_prob: float = 0.1, bias: bool = True):  
  
96         super().__init__()   
  
99         self.d_k = d_model // heads  
  
101        self.heads = heads  
  
104        self.query = PrepareForMultiHeadAttention(d_model, heads, self.d_k, bias=bias)  
105        self.key = PrepareForMultiHeadAttention(d_model, heads, self.d_k, bias=bias)  
106        self.value = PrepareForMultiHeadAttention(d_model, heads, self.d_k, bias=True)
```

Run the training pipeline

Next, you run the DAG to start the training job by invoking the `metl` parameters:

- `dataset`: The dataset resource to train the model.
- `model_display_name`: The human readable name for the trained model.
- `training_fraction_split`: The percentage of the dataset to use for training.
- `test_fraction_split`: The percentage of the dataset to use for testing.
- `validation_fraction_split`: The percentage of the dataset to use for validation.
- `budget_milli_node_hours`: (optional) Maximum training time in millihours (1000 = hour).
- `disable_early_stopping`: If `True`, the entire budget is used. Else, the training is completed before using the entire budget if the service believes it can improve on the model objective measurements.

The `run` method when completed returns the model resource.

The execution of the training pipeline will take upto 60 minutes.

```
[ ] model = dag.run(  
    dataset=dataset,  
    model_display_name="salads",  
    training_fraction_split=0.8,  
    validation_fraction_split=0.1,  
    test_fraction_split=0.1,  
    budget_milli_node_hours=20000,  
    disable_early_stopping=False,  
)
```

小结：高能人工智能平台HepAI



AI平台不是纯硬件平台，它本身是软件系统，承载AI算法模型，打通数据通道，提供AI算力。

AI
共
性
支
撑
能
力

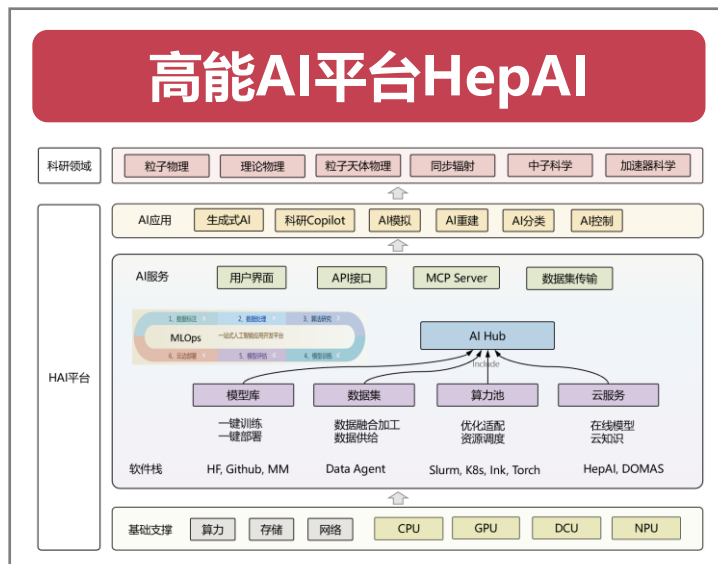
01 基础设施（算力）

02 大模型

03 科学软件&工具

04 AI-Ready数据集

整合



🚀 平台价值

- 对内提供公共服务，降低AI门槛
- 对外形成交流渠道，提升影响力

(1) 公共模型服务

- 大语言模型：DeepSeek, etc
- AI算法模型：SAM, ParT etc
- 科研工具：PDF解析 etc
- 一键部署、一键训练

(2) AI应用服务

- 生成式AI：HaiChat
- 学术助手：HaiAcademic(姜)
- 论文写作助手：Overleaf(侯)
- 快速开发/集成

(3) 公共算力服务

- 5PFlops , 200T SSD
- 领域模型在国产设备的适配
- AI易用（像私有服务器一样）
- 高效（智能调度提升整体利用率）

(4) 公共数据集服务



- AI4Data & Data4AI → 科学数据体系
- AI&ML for HEP → 大小模型协同
- AI Agent → 迈向AI驱动的自主科学发现
- HepAI → 共性基础设施



Thanks for Listening

02 AI共性支撑：AI模型、软件&工具服务



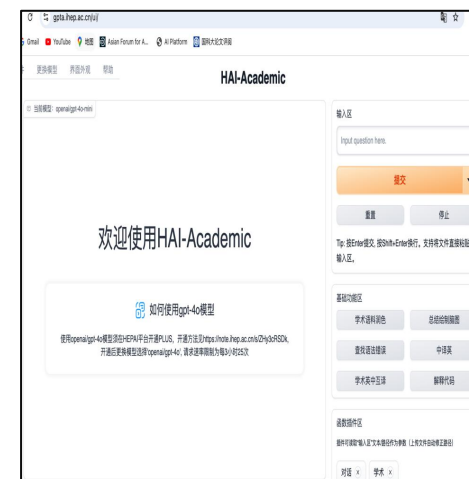
基于HepAI提供便捷易用的领域人工智能服务

- **大语言模型**：DeepSeek, ChatGPT, Xiwu, LLaMA3, Qwen etc
- **科学软件**：BOSS, Daisy
- **AI算法模型**：SAM(分割一切), PtychoNN(叠层成像重建)、ParticleNet&ParT(粒子鉴别), PointNet (事例重建)
- **科研工具**：PDF解析, PDF翻译, 论文润色, 科研绘图, etc
- **AI 应用**:
 - ✓ 生成式人工智能服务HaiChat
 - ✓ 学术论文助手HaiAcademic (姜法宇)
 - ✓ 论文写作助手Overleaf Copilot (侯丰尧)
 - ✓ 一站式天文警报汇集平台 (王文帅)
- **Zotero助手、VSCode代码补全**

HaiChat-v3界面：



HaiAcademic界面：

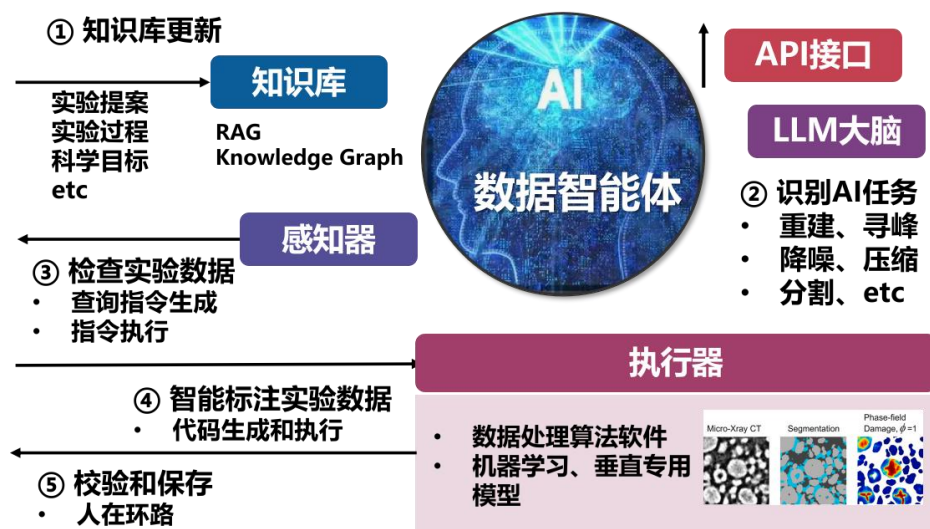


未来计划：涉及模型、软件、数据集、工具、应用等更多的服务，打造成AI+高能物理领域生态。



□ 基于智能体的AI-Ready数据标注

- ◆ 自动化程度高
- ◆ 效率高，初始精度可能低
- ◆ 自我改进（持续学习和反馈机制不断改进标注精度）
- ◆ 适合学科交叉（可提供数据信息理解和辅助）



□ 基于标注工具的AI-Ready数据标注

- ◆ 人工参与度高
- ◆ 高精度
- ◆ 灵活性强
- ◆ 适合领域专家



Local deployment of General-purpose Models



Local deployment of General-purpose Models on **GPU**, **NPU**, **DCU**

Generative Language Models

- ❑ DeepSeek-R1-671B.
- ❑ DeepSeek-R1-Distill-Qwen-32B
- ❑ Qwen3-Coder



Agentic Language Models

- ❑ xLAM-2-70b-fc-r
- ❑ xLAM-2-32b-fc-r



Embedding & Reranking Models

- ❑ BGE-M3
- ❑ BGE-Reranker-V2-M3



Specialization

- ❑ Text generation
- ❑ Code generation
- ❑ Dialogue systems
- ❑ Reasoning tasks

- ❑ Multi-turn dialogue
- ❑ Function calling

- ❑ Text representation,
- ❑ Information retrieval
- ❑ Relevance ranking

AI Activities in IHEP (Since 2022)



2022年

2022.09 第一届机器学习x高能物理研讨会

高能所机器学组Innovation Group和Production Group成立

2022.12 所创新：支持机器学习x高能物理、物理引导的机器学习



高能所战略研讨会

2023年

2023.06 高能所战略研讨会 (AI4HEP, 张正德)

2023.07 所“从0到1”：支持大模型x粒子物理 (Dr.Sai)

2023.08 高能物理人工智能平台HepAI上线。

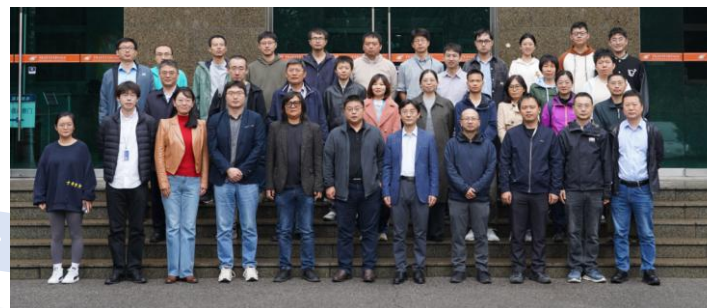
2023.11 中法粒子物理实验室, 2024FCPPL

2024年

2024.02 高能所全所AI4S规划会议, AI算力平台种子基金。

2024.05 高能所战略研讨会 (ML&HEP, 李科)

2024.10 高能所机器学习合作组成立



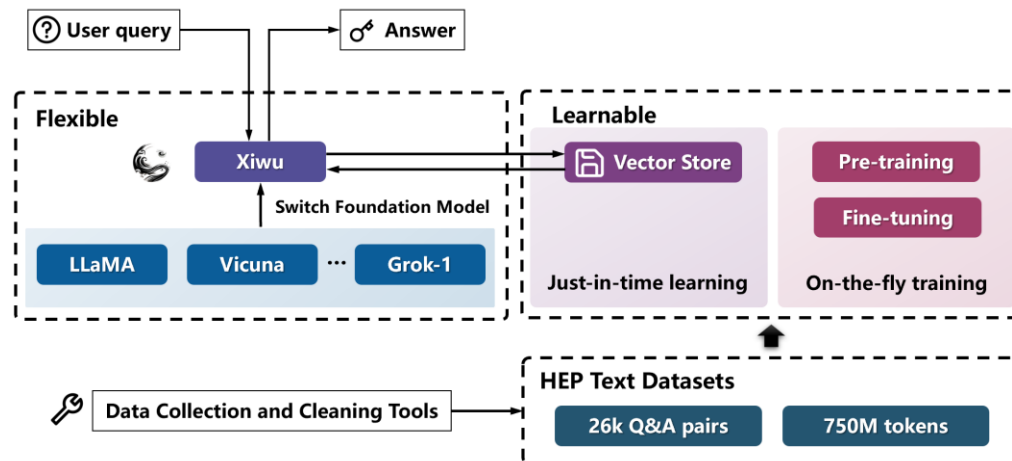
高能所机器学习合作组成立 (10.16)

Enhance Domain Capabilities

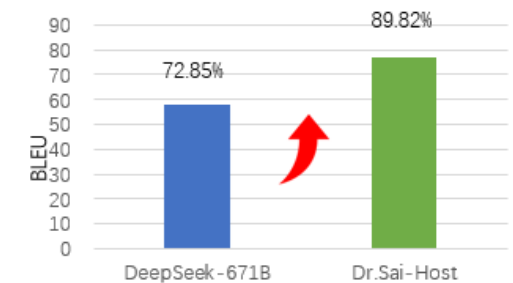


Brain **Xiwu(溪悟)** : A Basis Flexible and Learnable LLM for High Energy Physics ¹

- ❑ The **Brain** of Dr.Sai Agent.
- ❑ Currently, Xiwu is based on **DeepSeek-R1-Distill-Llama-70B**, Historically, LLaMA1~3.
- ❑ Secondary pre-training and fine-tuning techniques.
- ❑ Trained on GPU-A100 previously, now trained on DCU-K100 (Domestic GPU)
- ❑ Significantly better than the base model in HEP Q&A and internal code generation.
- ❑ New Model: **xLAM-2-32b-fc-r**, for function calling

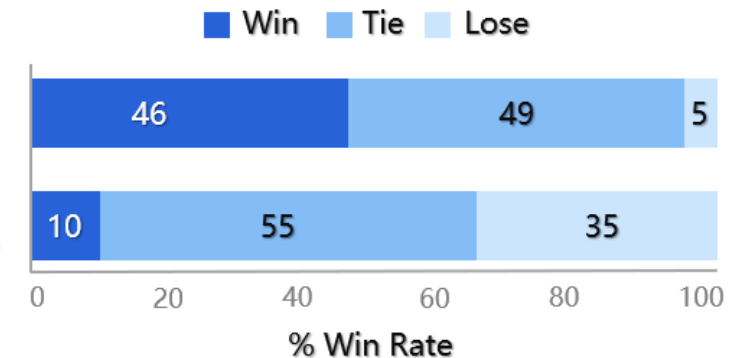


Test Results



Xiwu-13b
vs. Vicuna-13b

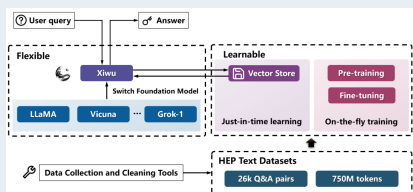
Xiwu-13b
vs. ChatGPT-175b



¹ [arXiv:2404.08001](https://arxiv.org/abs/2404.08001)



Brain



HEP·Xiwu LLM

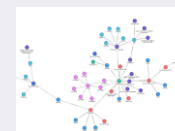
arXiv:2404.08001

- Based on LLaMA3-8B, 70B, and DeepSeek
- Trained on A800 and domestic DCU
- Enhanced domain capabilities

Memory



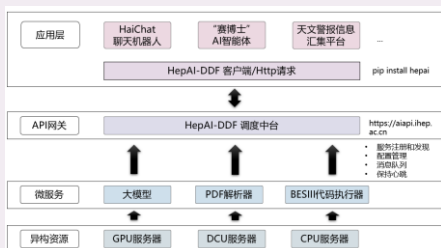
Vector Store



Knowledge Graph

- Vector Store
 - BAAI BGE-M3
 - LLaMA Index
- Knowledge Graph
 - LightRAG

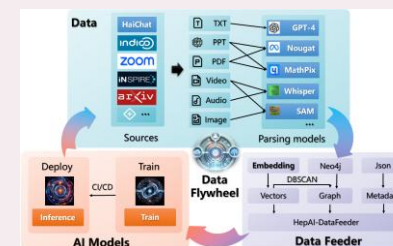
Actuators



HepAI DDF v2.0

- BESIII Actuator
 - BOSS Worker
- Daisy Framework
 - HEPSC
 - HEPSPtycho etc
- Unified access of models, tools etc

Sensors



Data Flywheel System

- Data Flywheel
 - HaiNougat
 - HaiChat
 - Whisper, SAM
 - ArXiv, Indico
- Domas Framework
 - Scientific metadata

成果水平、创新性



- HEP·Xiwu是高能物理领域领域定制化的大语言模型，利用了先进的LLM技术，实现了更加准确的领域文本处理。
- Dr.Sai是首个公开发布的高能物理科研智能体，相比与Cern提出的chATLAS, AccGPT等项目，它不仅能提供领域知识问答，而且能自主规划和执行操作，在设计和实现上具有先进性。

chATLAS

An AI Assistant for the ATLAS Collaboration

DANIEL MURNANE, GABRIEL FACINI,
RUNZE LI, DANIELE DEL SANTO & CARY RANDAZZO



Data:

	DB	Twiki	ATLAS Software Docs	E-group Archive	Indico Meetings	PDF Plots	Mattermost	Jira	ATLAS Codebases	Group level Docs	CDS Papers & Notes
Task Scrape		2k+	500+	10k+	~1k+	~1k+					~5k+
Convert											
Chunk & Embed											

Model: GPT-3.5 & GPT-4

AccGPT

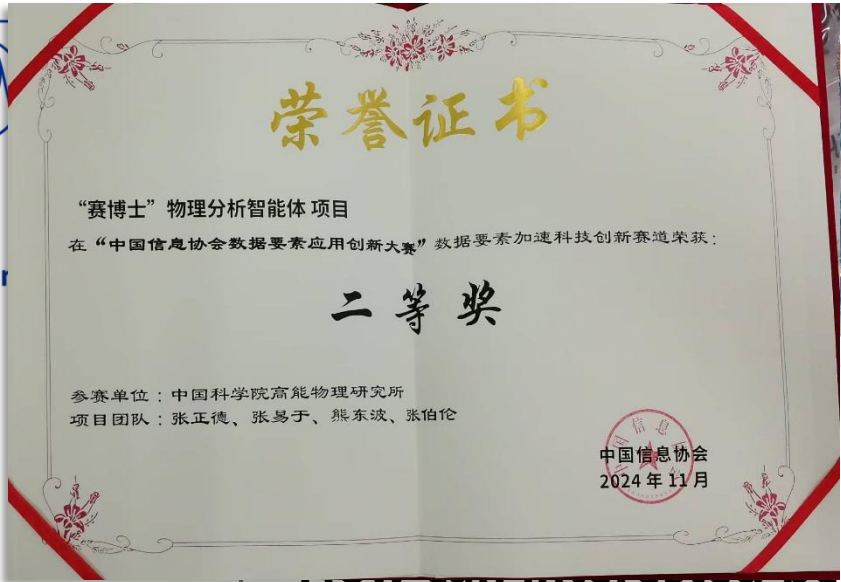
The Current Vision for AI Assistant
Accelerator Control and Beyond

Florian Rehm (CERN - Beams Department)

Idea:

1. Hysteresis compensation AI
2. Automatic and dynamic beam scheduling
3. Automatic LHC filling AI
4. Auto-pilots AI
5. Automatic fault analysis, recovery and prevention AI
6. Automatic testing and sequencing
7. Automatic parameter optimization AI

Model: LLaMA-13B

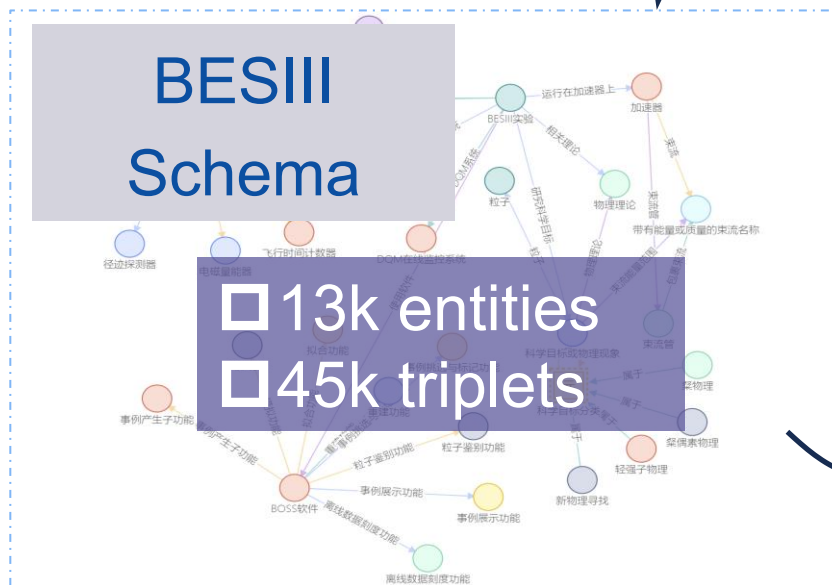


without expert knowledge
analyse CERN collision
data?
CERN Open Data
GPT-3.5 & GPT-4, Plugins

BESIII Knowledge Graph

- ❑ Knowledge Graph (KG-RAG)
 - ❑ Knowledge Extraction (BESIII Schema)
 - ❑ Knowledge Mapping (Entity Alignment, Relationship Chaining)
 - ❑ Graph Storage (Neo4j)
 - ❑ KG-RAG system (lightweight RAG)

**600
BESIII
papers**



Self-reflection based on mapping



Tell me the information about Y(4360)



[Incorrect Output] $Y(4360)$ is ...

Verify

[Verified Output] Y(4360) is ...

[Corrected Output] Y(4360) is ...

Correct

Actuator

HepAI-DDF

High-performance AI Distributed Deployment Framework

Core Features

- ✓ **Unified Access:** Seamless integration of AI models, scientific tools, vector databases, and knowledge graphs across heterogeneous resources (CPU/GPU/DCU).
- ✓ **Smart Scheduling:** Automated heartbeat monitoring + dynamic load balancing.
- ✓ **Unified API Gateway :** Standardized service interfaces with API-KEY authentication for enhanced security.
- 🔥 **Remote Model Invocation :** NEW Groundbreaking support for remote server model deployment with seamless local calling experience.
- 🔥 **Adapted MCP:** Model Context Protocol, Access to 2600+ MCP Server.



- ❑ 50+ Models
- ❑ 2 local tools
- ❑ N MCP tools
- ❑ In one server

❑ HaiDDF2.0 was lanced in Feb. 2025.

AI Agent for Diffraction Data (Simu. & Exper.)



Constructing a Cross-Instrument X-ray and Neutron Powder Diffraction Experimental and Simulated AI Dataset

◆ Example of Rietveld Refinement :

Input: Experimental X-ray/neutron diffraction data, Instrument parameters, Initial crystal structure (CIF)

gsasii_xrd_process_agent发言:

已完成2次晶胞参数精修, 结果如下:

精修质量指标:

1. X射线数据 (PWDR PBSO4.XRA Bank 1) :
 - Rwp值: 14.009 (较初始值降低约12%)
2. 中子数据 (PWDR PBSO4.CWN Bank 1) :
 - Rwp值: 4.715 (达到高精度收敛)

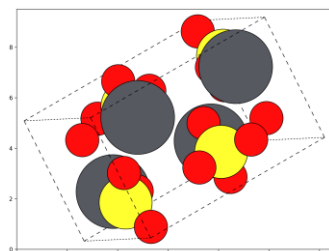
精修后晶胞参数 (Phase_1) :

参数	长度 (Å)/角度 (°)	物理意义
a轴	8.4718	晶格沿x方向基本矢量
b轴	5.3932	晶格沿y方向基本矢量
c轴	6.9538	晶格沿z方向基本矢量
体积	317.72 Å³	单胞空间占位量
晶系	正交晶系	$\alpha=\beta=\gamma=90^\circ$

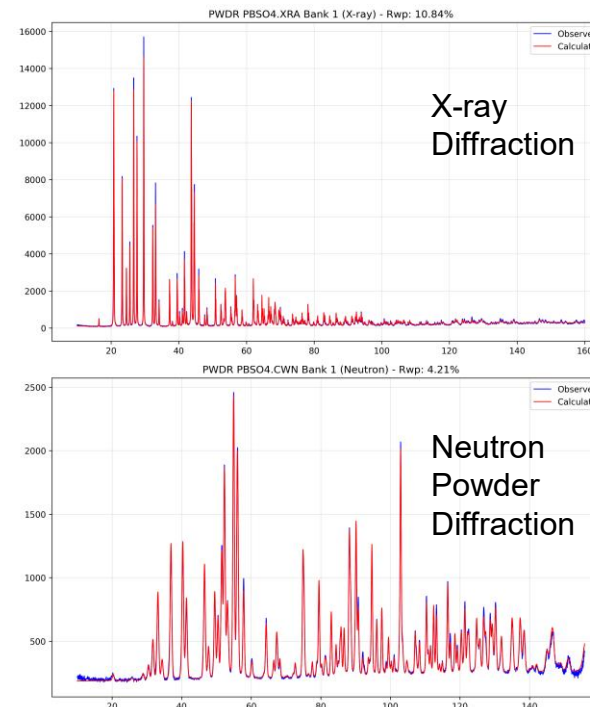
需要我为您:

1. 生成GSAS-II精修报告 (PDF格式)
2. 绘制XRD实验-计算拟合对比图
3. 输出精修后CIF文件
4. 继续执行其他参数精修 (如原子位置/温度因子)

TERMINATE



Output: Refined crystal structure, Simulated diffraction patterns matching experimental data



□ **AI-ready Dataset:** Crystal structures (initial & refined), Simulated X-ray/neutron powder diffraction patterns, Experimental X-ray/neutron powder diffraction data.

✓ Completed refinement for selected crystal structures using combined experimental/simulated data.

➤ Currently expanding dataset by collecting cross-instrument experimental data (X-ray & neutron) from battery material samples for AI applications.



Construction of a **Pair Distribution Function (PDF)** Experimental-Simulation Fusion Dataset

◆ PDF Refinement Example:

① Input: PDF data file + structural file

我需要数据ID 878d02e0-e521-44dd-aaf3-512d4b61dfa6中的Ni样品X-ray PDF数据进行精修，精修参数为：Q最大值 (1/Å) 为30.0，Q分辨率因子为0.01，精修最小值 (Å) 为1.5，精修最大值 (Å) 为20。处理结果储存在数据空间dc801e64bcab42cf80ee0569db0fa9f1中。

SCI-DataAgent
正在处理数据集878d02e0-e521-44dd-aaf3-512d4b61dfa6...
正在总结数据集信息，请稍后...
思考用时 19 秒

摘要：
本数据集通过diffpy.pdf2软件对Ni样品的X射线PDF数据进行精修，精修参数包括Q最大值30.0 (1/Å)、Q分辨率因子0.01、精修范围1.5-20 Å。最终精修结果表明，晶格参数为3.5316 Å，R因子 (Rw) 为7.00%，比例因子 (pscale) 为0.732，温度因子为0.0054，系统误差参数 (qdamp) 为0.063，位移参数 (delta2) 为2.253。结果显示拟合精度较高，能够有效反映Ni晶体的局部原子结构特征。

****卡card:**
数据集

研究领域：
X射线PDF分析/X-ray PDF analysis, 对分布函数/Pair Distribution Function, 晶体结构精修/Crystal Structure Refinement正在上传数据集到数据空间中...

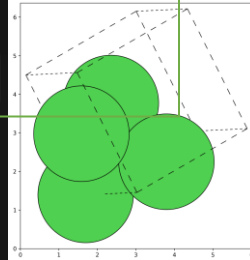
数据集Ni-xray_pdf_refined成功上传到数据空间dc801e64bcab42cf80ee0569db0fa9f1中!

没有

摘要：
本数据集通过diffpy.pdf2软件对Ni样品的X射线PDF数据进行精修，精修参数包括Q最大值30.0 (1/Å)、Q分辨率因子0.01、精修范围1.5-20 Å。最终精修结果表明，晶格参数为3.5316 Å，R因子 (Rw) 为7.00%，比例因子 (pscale) 为0.732，温度因子为0.0054，系统误差参数 (qdamp) 为0.063，位移参数 (delta2) 为2.253。结果显示拟合精度较高，能够有效反映Ni晶体的局部原子结构特征。

卡片：

数据集包含Ni样品的X射线PDF精修结果文件 (.pdf, .rstr, .res) 及可视化图谱 (.png)，总数据量约517 KB。数据来源为实验采集的Ni



② Processing & Task Allocation

③ Output: Saved result paths/links

数据文件，共有4个文件，目前只展示前10个文件。:

文件名	PDF	网址	文件路径
Ni-xray.fgr文件	PDF精修结果文件	https://alip1.hep.ac.cn/api/v2/files/file-3e4daf3c39/preview	/afs/user/home/xiongb/datasets/dataagent/test_datasets/Ni_pdf_test/refined_llqm@hep.ac.cn/Ni-xray.fgr
Ni-xray.rstr	PDF精修结果文件	https://alip1.hep.ac.cn/api/v2/files/file-a3ee78e1a1/preview	/afs/user/home/xiongb/datasets/dataagent/test_datasets/Ni_pdf_test/refined_llqm@hep.ac.cn/Ni-xray.rstr
Ni-xray.res	PDF精修结果文件	https://alip1.hep.ac.cn/api/v2/files/file-fbc811095d/preview	/afs/user/home/xiongb/datasets/dataagent/test_datasets/Ni_pdf_test/refined_llqm@hep.ac.cn/Ni-xray.res
Ni-xray.png	PDF精修结果文件	https://alip1.hep.ac.cn/api/v2/files/file-87a5118c20/preview	/afs/user/home/xiongb/datasets/dataagent/test_datasets/Ni_pdf_test/refined_llqm@hep.ac.cn/Ni-xray.png



Fig. PDF refinement results

□ **AI-ready Dataset:** Crystal structures, simulated PDF data, experimental PDF data.

- ✓ PDF refinement for sample datasets;
- currently validating refinement accuracy on additional experimental data.


```

graph TD
    1[① User Proposal] --> 2[② Natural Language Processing & Task Decomposition]
    2 --> 3[③ Experimental Events]
    2 --> 5[⑤ Extract Instrument Parameters]
    2 --> 6[⑥ Auto-Search Crystal Structure Files (via mp-id)]
    3 --> 4[④ Diffraction Pattern]
    3 --> 5
    3 --> 6
    4 --> 7[⑦ Pattern Refinement]
    7 --> 8[⑧ PDF Refinement]
    8 --> 9[Experimental Report]
    
```

④ Diffraction Pattern

BANK	5	1599	1599	RALF	29504	634	30784	0.020974	FXYE
922.491684	3.2184362444	0.0041362261							
932.413174	3.1877303973	0.0040709853							
952.216963	3.1740813673	0.0040169919							
971.944276	3.1504766725	0.0039594780							
991.598796	3.1440673305	0.0039128785							
1011.183973	3.1296857705	0.0038639496							
1030.703045	3.1062754701	0.0038140750							

⑤ Extract Instrument Parameters

```

ZEPOL: 58 - 26742764808784
difc: 10265.470951206315
difA: -2.773782385159114
difB: -14.239510067597374
alpha: 0.016700628080595446
beta: 0.0.02418965394654335
beta-1: 0.001200275767666099
beta-q: -0.00038442372919489305
sig-0: -447.3963235002162
sig-1: 1212.7105730694173
sig-2: 108.199523242157919
sig-q: 561.7692176934811
  
```

⑥ Auto-Search Crystal Structure Files (via mp-id)

```

[FunctionCall(id="call_bfxte7vxt0abshs4ubu5qy05", arguments={"material_ids": ["mp-149", "mp-13"]}, name="get_material_data")]
Retrieving SummaryDoc documents:
100%|
----- did -----
[FunctionExecutionResult(content=[TextContent(type="text", text=["status": "success", "count": 2, "materials": [{"material_id": "mp-13", "formula": "Fe", "structure": [{"@module": "pymatgen.core.structure", "@class": "Structure", "charge": 0, "lattice": [{"matrix": [{"2.33614509, 0.00011167, -0.8258229300000001}], [{"1.16807798, 2.02304724, -0.8260822}], [{"1.17007387, 2.027305, 3.31032796}], "pbc": [true, true, true], "a": 2.477813026804854, "b": 2.477809133849768, "c": 4.054295207189474, "alpha": 90.00003845025343, "beta": 90.00031950636807, "gamma": 109.469836683298, "volume": 23.468168469356993}], "properties": [{"sites": [{"species": ["Fe", "occu": 1]}, "abc": [{"0.50000002, 0.50000008, 0.49999999}], "properties": [{"magmom": 2.758, "label": "Fe", "xyz": [{"1.1690704315759348, 2.025232096572963, 0.8292112992936858}], [{"species": [{"element": "Fe", "occu": 1}], "abc": [{"0.99999998, 0.99999992, 1e-08}], "properties": [{"magmom": 2.758, "label": "Fe", "xyz": [{"1.1680671684240749, 2.023158768427037, -1.6519050142936857}]}], "origins": [{"name": "structure", "task_id": "mp-2739258", "last_updated": "2021-07-24T19:15:26.927000"}], {"name": "energy", "task_id": "mp-2298433", "last_updated": "2025-02-
  
```

Experimental Report

User Name:
Experiment Date:
Sample 1: Chemical Formula: Run ID:
Experimental Data:
=====

GSAS Refinement Results:
PDF Refinement Results:
Space Group:
Lattice Parameters:
3D Crystal Structure:
=====

Sample 2: [...]

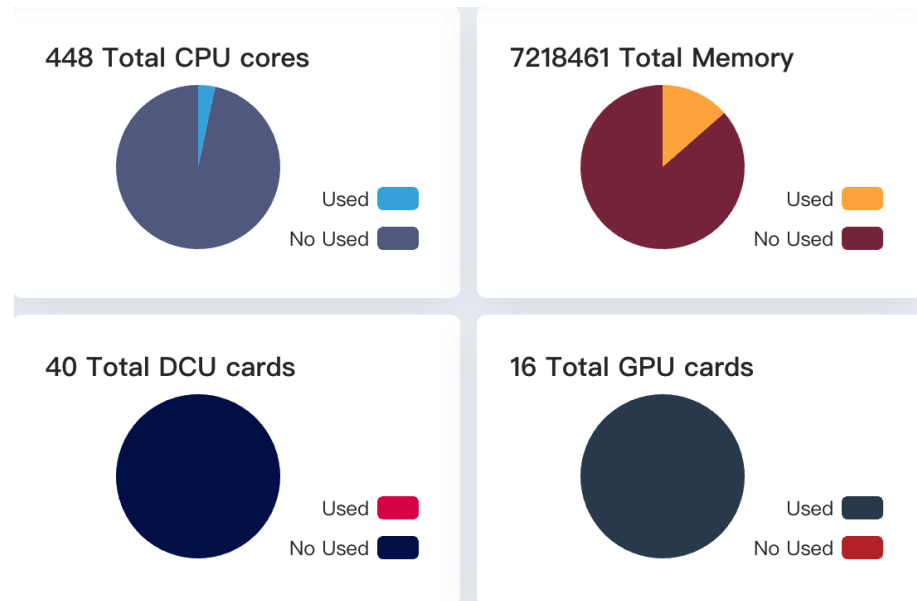


■ 异构资源

- GPU: 1*8卡L40 + 1*8卡A800
- DCU: 6*8卡K100-AI
- NPU: 2*8卡910B

■ 亮点及效果

- 国产 DCU 加速卡首次在高能物理领域成功运用
- AI 大模型 (8B 70B) 在 DCU 上的部署优化: 单卡DCU K100-AI 性能从 A100 的 25% 提升到 45%
- AI 集群实现了本地 DeepSeek 模型的部署
 - deepseek-r1:32b
 - deepseek-r1:671B (满血版)
- 完成 DCU 适配模型 Particle Transformer, Xiwu, ParticleNet, ResNet-CMAB, LLaMA



启动服务中请耐心等待!

已用时间: 00:23 ————— 预计剩余时间: 11:37

AI Model for Event-to-spectrum Prediction

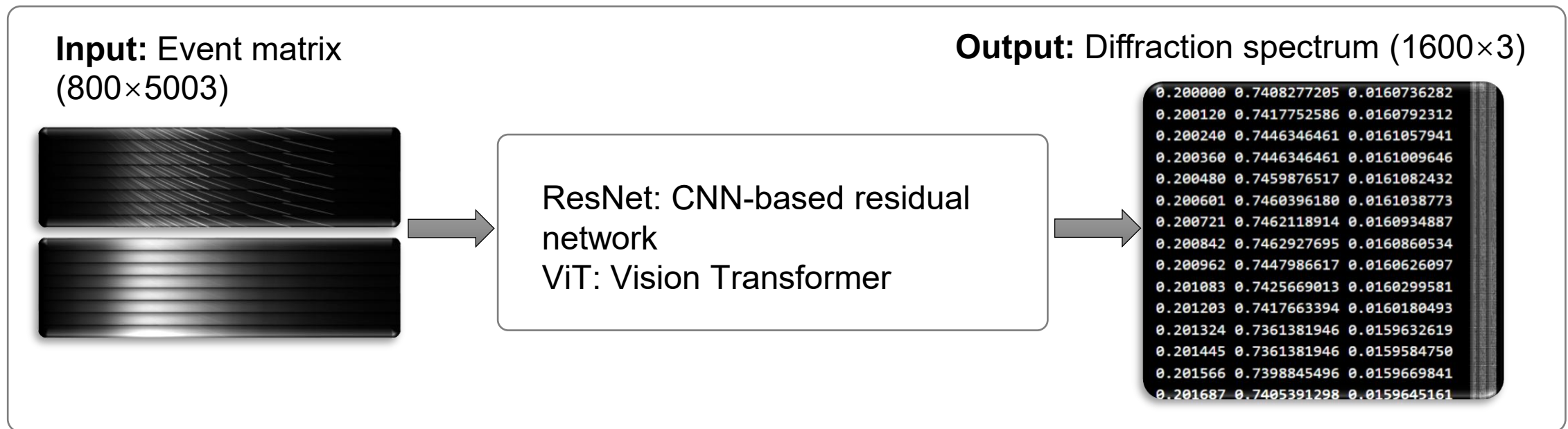


Hong WANG
Rong DU

NN

Event-to-spectrum prediction for neutron powder diffraction event data.

- ❑ Background: While **Rongzai** software can process neutron powder diffraction event data into diffraction spectrum, users face significant latency during data processing. This project develops an event-to-spectrum prediction **AI model** for instantaneous online diffraction visualization.
- ❑ AI Task: Build a combined simulated/experimental dataset to train a high-speed, high-accuracy event-to-pattern prediction model.



- ✓ Simulated data generation completed; ✓ Data preprocessing pipeline developed
- ✓ Model architectures modified for diffraction physics; ✓ Initial training and code validation performed