National Astronomical Data Center
国家天文科学数据中心

MCP和NADC智能体服务体系

许允飞代表VO团队

国家天文科学数据中心

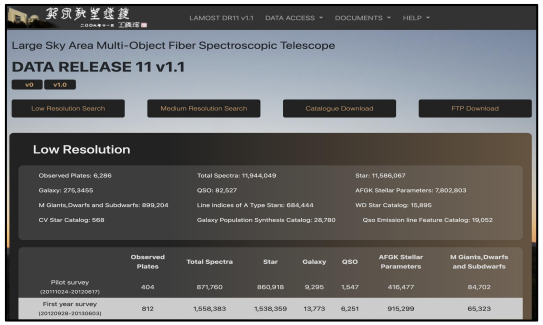
中国科学院国家天文台

2025-11-28

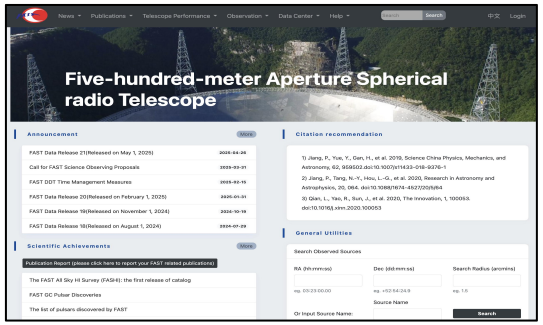
研究背景

National Astronomical Data Center
国家天文科学数据中心

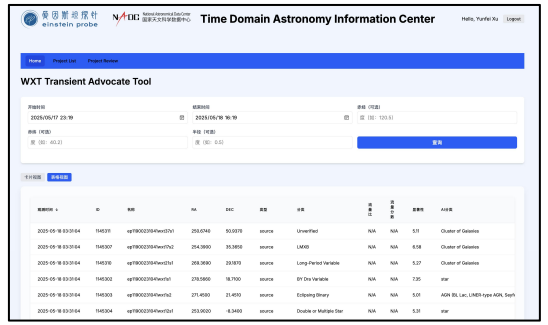
- 国家天文科学数据中心（NADC）经过多年建设，在国家重点研发计划、国家自然基金委重点项目、B类先导专项的支持下，与LAMOST、FAST、司天工程、天关卫星、CSST等重大科研任务深入合作，已经初步建成了一系列有效支撑天文科学研究的数据融合应用与服务。



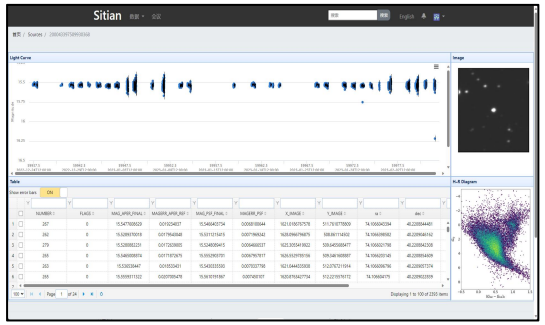
LAMOST数据发布系统



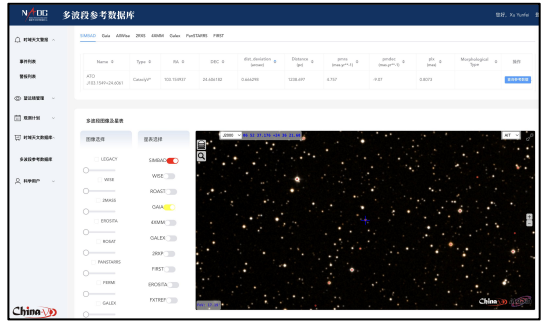
FAST在线信息系统



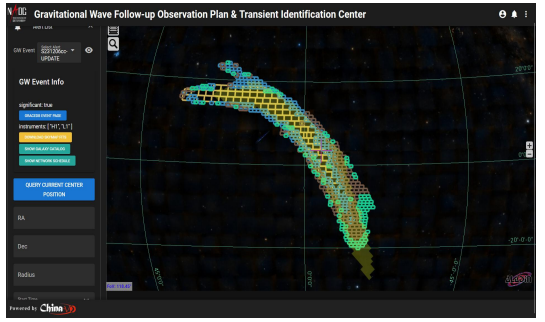
天关卫星时域天文信息中心



mini司天数据发布系统



天地一体化观测-数据-研究协同平台



引力波后随观测规划系统

交互界面复杂，用户学习成本高

服务分散在多个系统中，缺乏协同

功能高度定制化，不够智能化

研究背景

National Astronomical Data Center
国家天文科学数据中心

- 在AI时代，如何借助LLM的能力，提供更便捷的数据服务，支持天文科研，促进重大成果产出。



“被关在地下室的博学家”
知识渊博但无法触及外部世界

知识滞后

不知道最新爆发的GRB事件， 训练数据有截止期。

幻觉

可能一本正经地编造不存在的观测数据。

无法行动

无法直接访问或处理天文数据。

MCP协议

National Astronomical Data Center
国家天文科学数据中心

- 需要以一种标准方式，将外部数据作为上下文（context）提供给大模型。
- MCP（Model Context Protocol）协议应运而生。
- 基于该协议，我们可以把 NADC 的数据服务封装成工具，安全地提供给任何支持 MCP 的大模型调用。

MCP (Model Context Protocol)

给大模型装上USB接口，即插即用



大模型

MCP 协议

标准接口



NADC 数据服务

MCP与API的关系

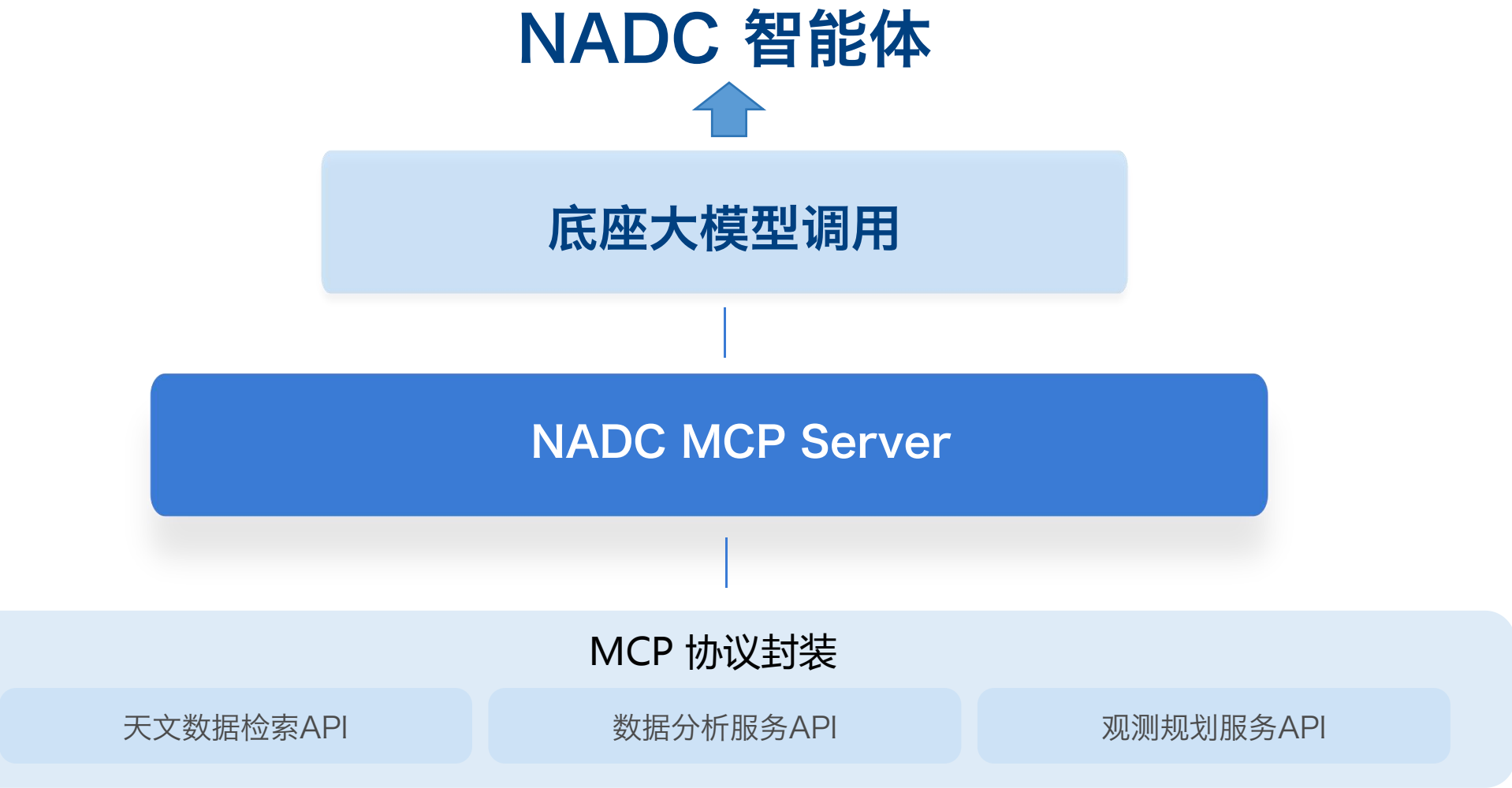
National Astronomical Data Center
国家天文科学数据中心

- API：
  - 每个服务都有自己独特的 API 定义
  - 实现的是点对点的、定制化的连接
  - 将N个API接入到M个大模型，需要N*M种集成

- MCP：
  - 标准化 AI 模型与数据源或工具之间的连接
  - AI 模型能够以统一的方式访问资源，无需单独编写特定的集成代码
  - MCP 是建立在 API 之上的一套高层协议

NADC MCP服务架构

National Astronomical Data Center
国家天文科学数据中心



智能体

- 智能体是一类**大模型应用**，它能够根据用户意图，自主决策并采取行动以实现目标的软件或硬件实体。

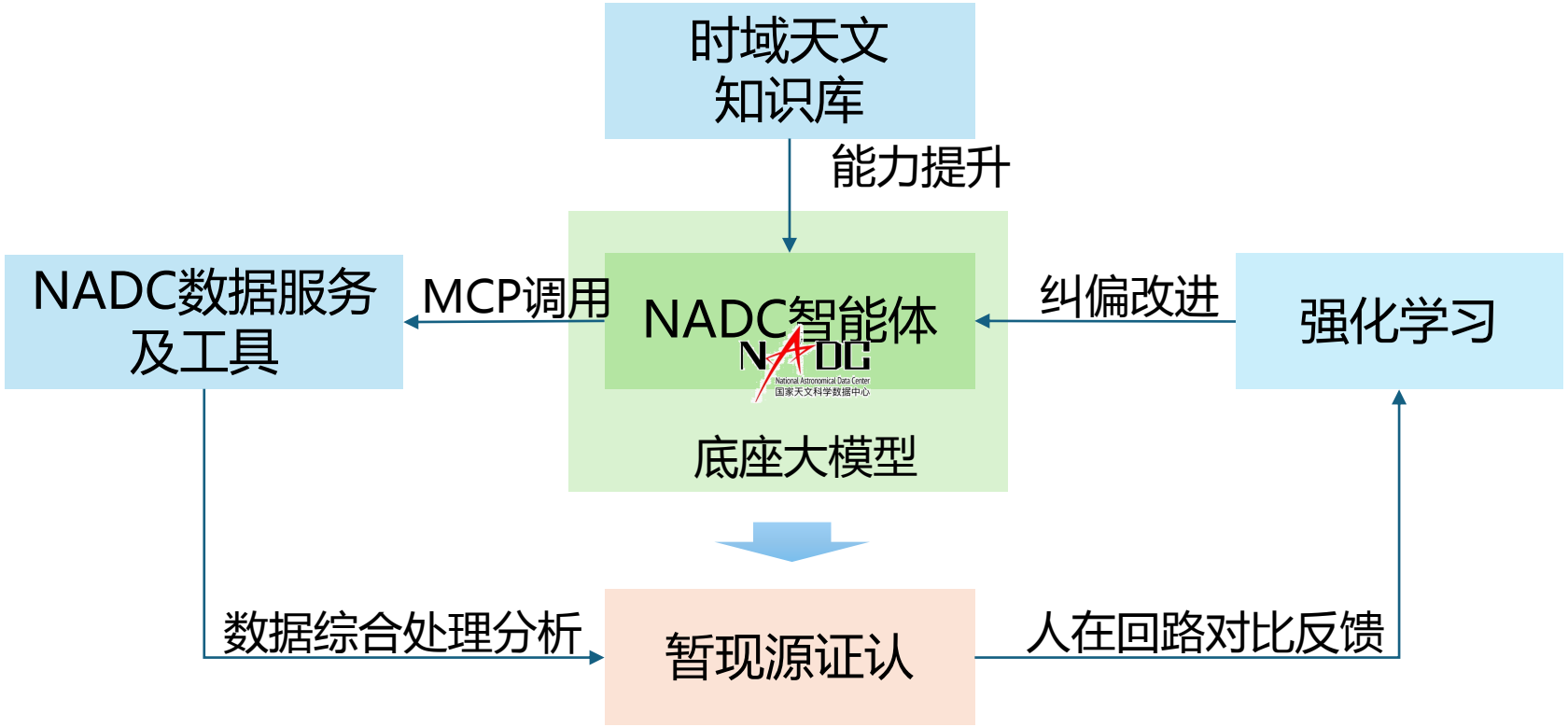
交互能力	与用户通过自然语言交互，理解用户意图，反馈执行结果
决策能力	根据用户意图自主决策，规划行动路径，并根据行动反馈修正决策
行动能力	能够调用外部工具，获取信息、改变状态，优化决策
学习能力	能够通过训练过程改进自身决策和行动策略，适应不断变化的需求

NADC智能体目标

National Astronomical Data Center
国家天文科学数据中心

近期目标：训练智能体使其具备时域天文知识，使其能够像经验丰富的科研人员一样自动检索多源数据、分析时域警报、处理暂现源数据，给出证认结论，并列出所有证认依据。

长远目标：辅助天文学家提出假设，自动设计研究方案，申请观测，获取并分析数据开展研究，得出科学结论。

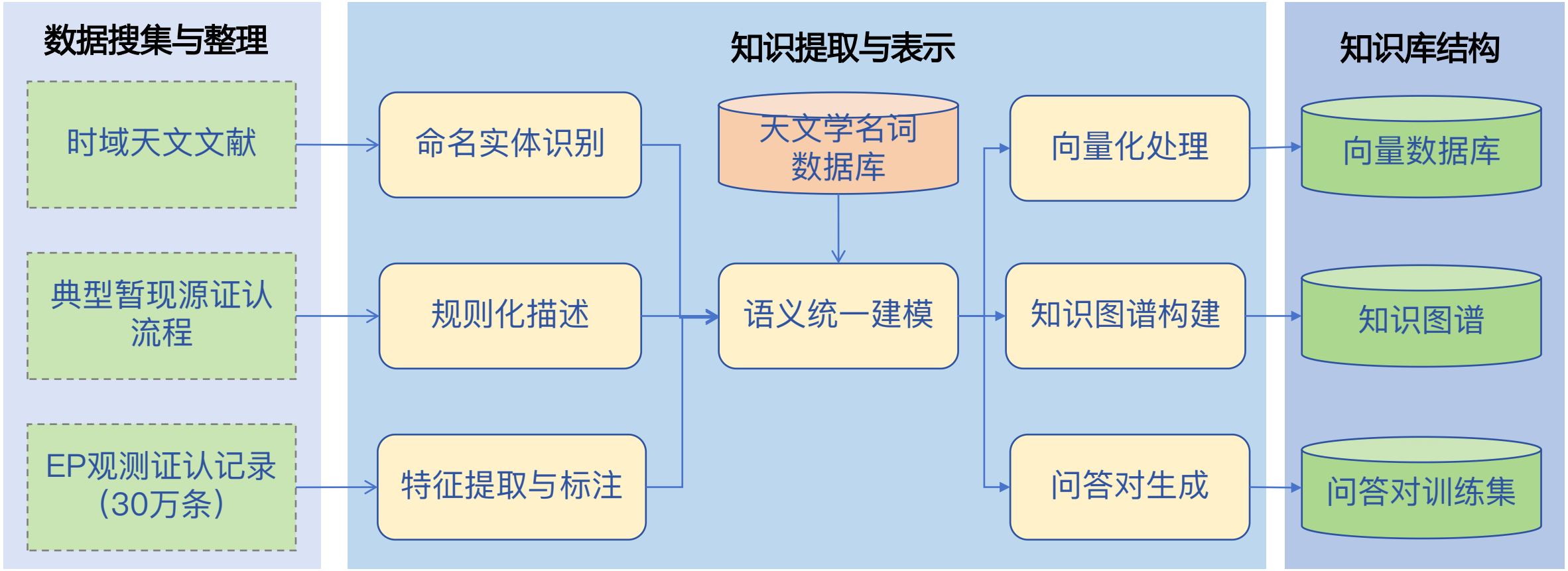


NADC智能体技术路线

智能体能力提升

National Astronomical Data Center
国家天文科学数据中心

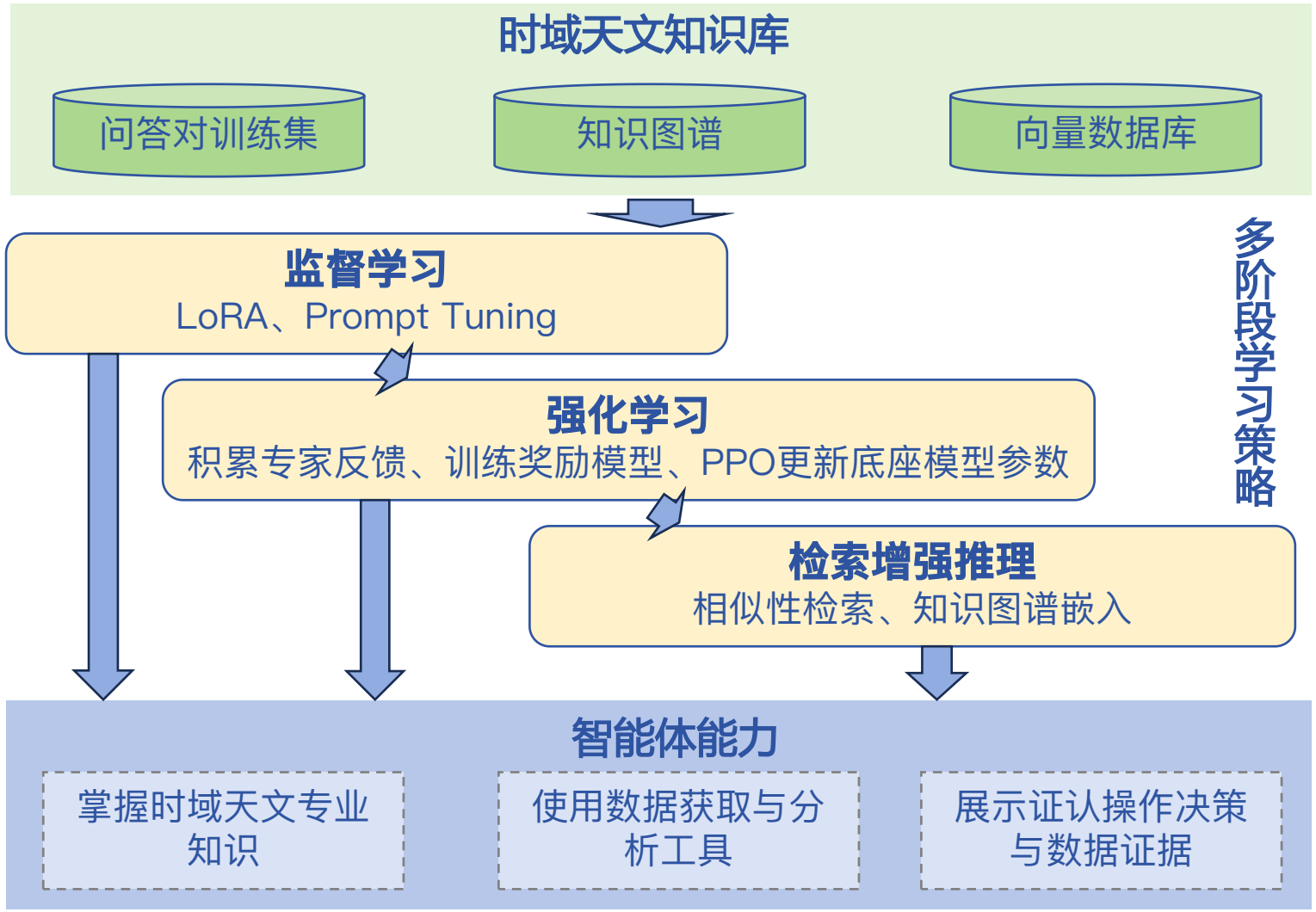
- 底座模型：AstroOne V1、Deepseek R1、Qwen3-32B（微调）、ScienceOne（磐石，中科院自动化所）
- 时域天文知识库：



智能体能力提升

National Astronomical Data Center
国家天文科学数据中心

- 多阶段学习策略：
  - LoRA
  - Prompt Tuning
  - 强化学习
  - GraphicRAG



智能体执行能力

National Astronomical Data Center
国家天文科学数据中心

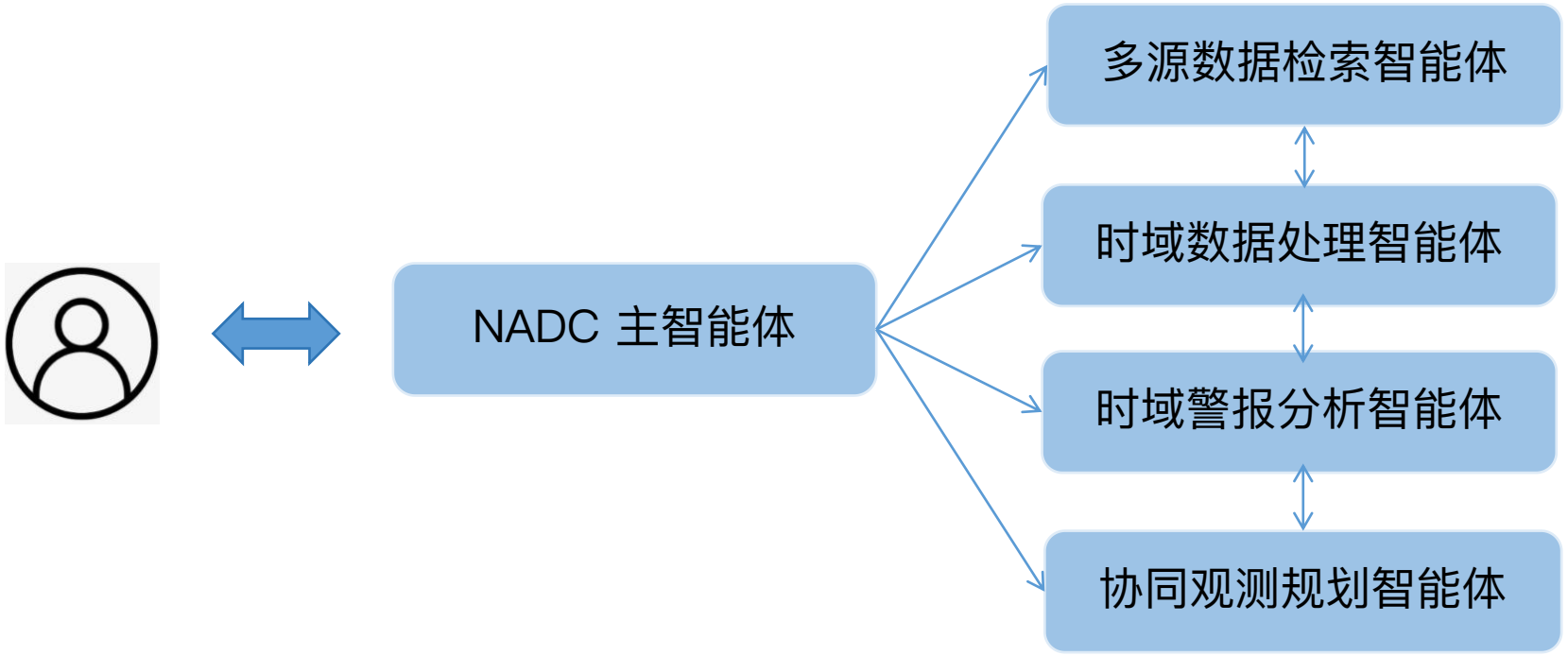
- 将NADC现有数据融合服务接入至大模型。智能体通过能力提升，知道何时应调用何种数据服务
- 服务接入方式：NADC MCP Server



智能体群协作

National Astronomical Data Center
国家天文科学数据中心

- 智能体-智能体群
  - 任务复杂性：单智能体接入外部工具过多，调用工具的准确率会急剧下降
  - 任务并行化：多个智能体可并行化的开展数据获取和数据分析
  - 能力拓展性：让每个智能体的能力专业化，可实现有针对性的优化提升



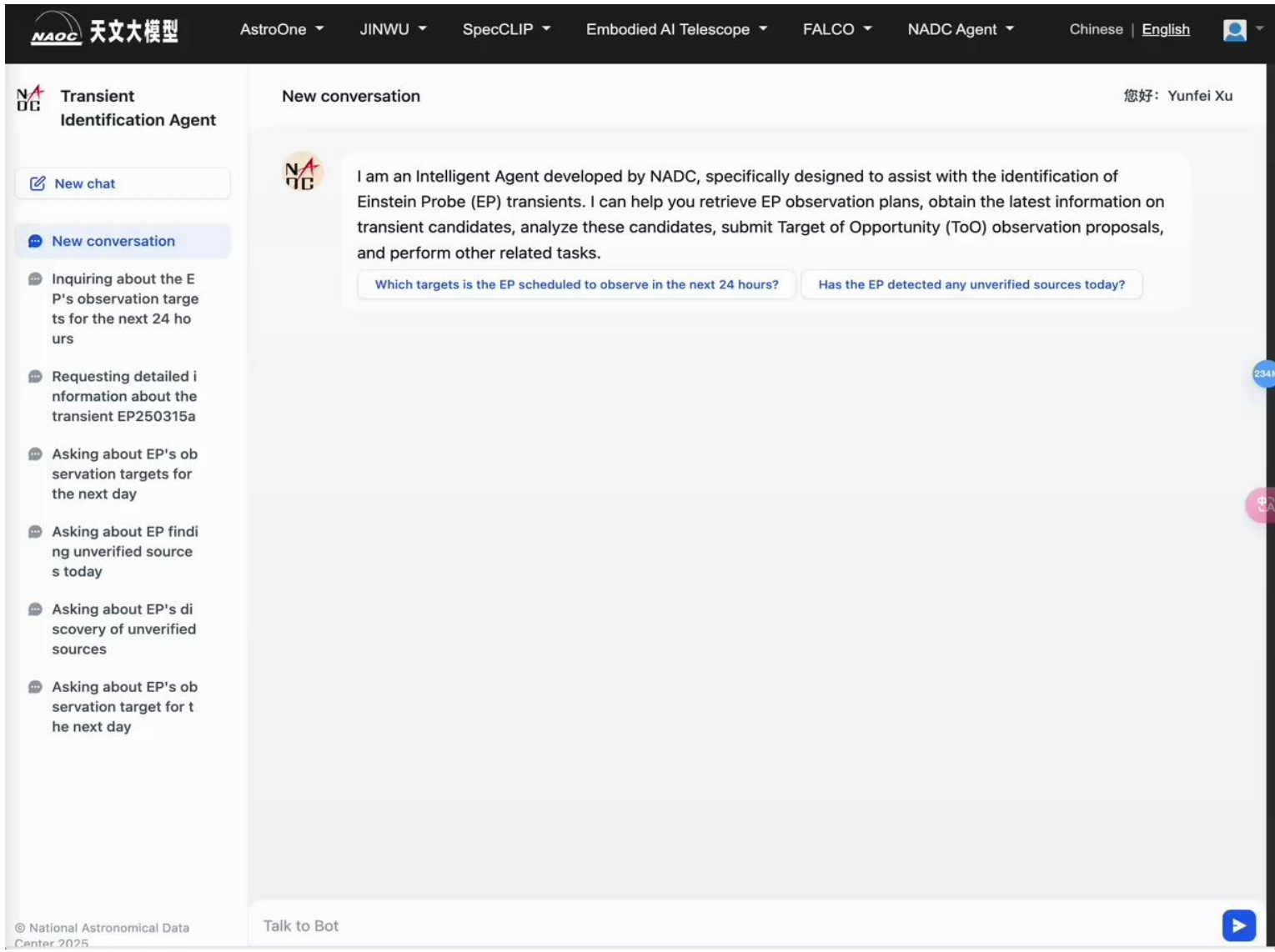
能力发现：每个智能体使用“名片”介绍自身能力和交互参数

标准化通讯：基于JSON-RPC2.0标准实现智能体间通讯

NADC智能体的典型应用

National Astronomical Data Center
国家天文科学数据中心

- 基于自然语言的数据查询
  - 列出数据详情
  - 给出原始数据出处
  - 多轮对话持续交互



NADC智能体的典型应用

National Astronomical Data Center
国家天文科学数据中心

EP暂现源证认

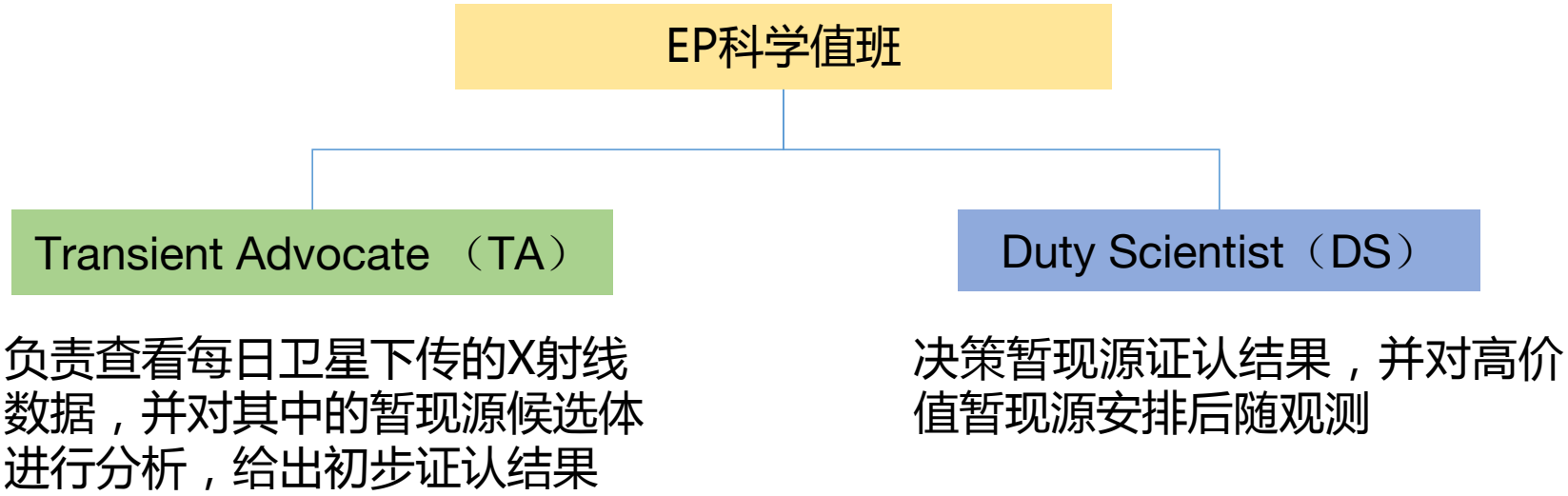
- 暂现源证认是EP卫星数据处理和分析最为核心的流程之一，也是EP科学发现的前提

1）从EP数据中探测到暂现源信号

2）对该信号进行判别及分类

3）对高价值暂现源进行地基后随观测

4）根据后随观测结果进一步判断暂现源性质



X射线源暂现源证认支持

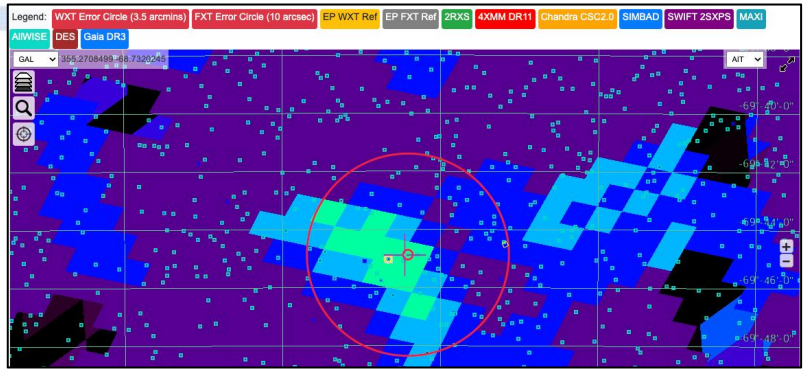
National Astronomical Data Center
国家天文科学数据中心

Candidate List

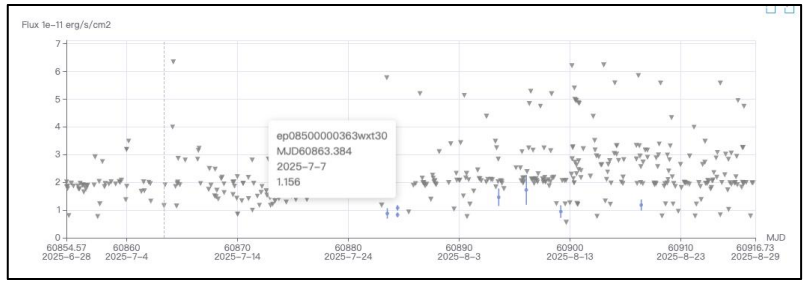
Export CSV

	Source ID	Obs Time (UTC)	Type	Tags	Category	Classification	Common Name	Submit	TA	Image	LC	RA	Dec	Esti_Flux (erg/cm^2/s)	Significance
			All												
1	ep01709244186wxt14s1	2025-09-20 00:49:33	source	known_source	AGN	BL Lac	[OKM2018] SWIFT J0244.8-5829		SY01Pipeline			41.189	-58.312	3.92e-11	10.388
2	ep01709244186wxt4s2	2025-09-20 00:35:45	source	known_source	AGN	Seyfert 2	LEDA 89420		SY01Pipeline			351.354	-38.458	1.46e-11	5.124
3	ep01709244186wxt4s1	2025-09-20 00:35:45	source	stellar_flare,known_source	Star	BY Dra Variable	WW PsA		SY01Pipeline			341.247	-33.255	2.08e-11	6.736
4	ep01709244186wxt1s1	2025-09-20 00:36:02	source	known_source	AGN	Seyfert 1	NGC 7213		SY01Pipeline			332.322	-47.175	2.9e-11	8.734
5	ep06800000909wxt48s1	2025-09-19 21:28:36	source	known_source	Unverified	Unverified	WXT J222847-600602		Fu Yucong			337.196	-60.101	1.38e-11	5.278
6	ep06800000909wxt40s1	2025-09-19 21:25:22	source	known_source	AGN	Seyfert 1	ESO 141-55		SY01Pipeline			290.313	-58.691	4.02e-11	9.722
7	ep06800000909wxt39s3	2025-09-19 21:25:14	source	known_source	Unverified	Unverified	WXT J203920-602124		Fu Yucong			309.833	-60.357	1.23e-11	5.091
8	ep06800000909wxt39s2	2025-09-19 21:25:14	source	known_source	Cluster of Galaxies	Cluster of Galaxies	ACO 3667		SY01Pipeline			303.133	-56.845	1.95e-11	5.18
9	ep06800000909wxt39s1	2025-09-19 21:25:14	source	stellar_flare,known_source	Star	High Proper Motion Star	BP Ind		SY01Pipeline			312.7	-56.239	5.85e-11	14.059
10	ep06800000909wxt37s3	2025-09-19 21:21:24	source	known_source	AGN	BL Lac	QSO B2005-489		SY01Pipeline			302.381	-48.781	1.44e-11	5.114
11	ep06800000909wxt37s2	2025-09-19 21:21:24	source	known_source	AGN	BL Lac	2FGL J1936.8-4721		SY01Pipeline			294.236	-47.346	8.81e-12	5.26
12	ep06800000909wxt37s1	2025-09-19 21:21:24	source	transient			EP250919a		SY01Pipeline			298.57	-48.861	1.23e-11	5.636
13	ep06800000909wxt22s1	2025-09-19 21:13:17	source	known_source	AGN	Seyfert 1	NGC 6814		SY01Pipeline			295.7	-10.351	1.4e-11	5.33

待证认暂现源候选体列表
高亮条目是流量相比历史观测有显著变化的候选体



假源剔除、多波段数据检索分析



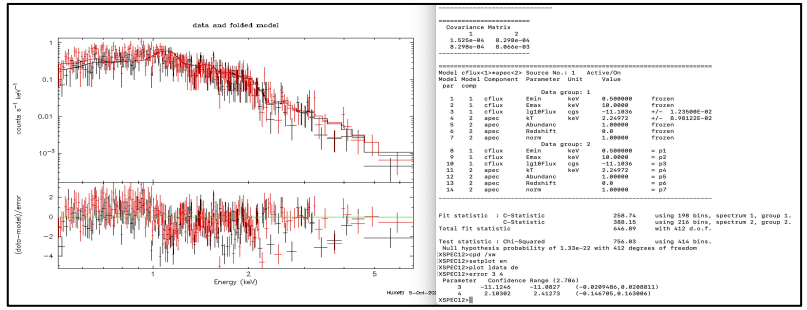
长期光变分析

TA需人工进行暂现源候选体分析。每个花费10-30分钟不等。

平均每天50个需人工检查的暂现源候选体，任务枯燥繁重



后随观测及数据分析



能谱拟合分析

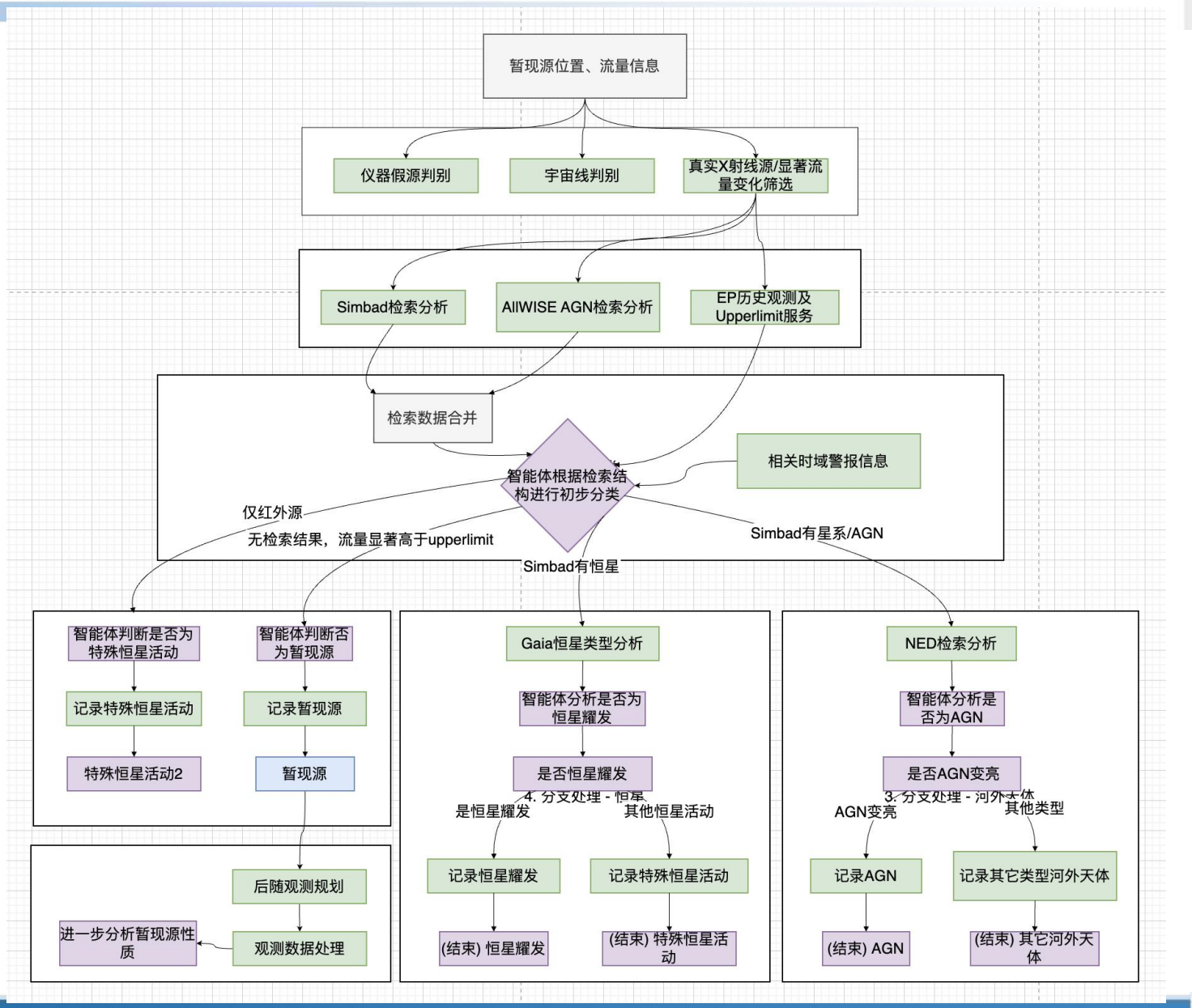
NADC智能体的典型应用

National Astronomical Data Center
国家天文科学数据中心

天关卫星暂现源智能化证认

- Qwen3-32B经过微调及强化学习后，对于EP暂现源候选体的典型证认流程

- 绿色模块是智能体的工具调用
- 紫色模块是智能体的思考过程



应用效果

- 智能体实时处理每日下传数据，分析其中提取的X射线源
- 证认报告将展示在TA日常证认使用的源数据页面上，包括证认结论和依据数据
- TA可对证认报告打分，并给出文字反馈
- TA可与智能体进一步对话，提问或质疑

类别	证认效率	准确率	召回率
宇宙线	~100ms	0.993	0.992
十字臂假源	~300ms	0.975	0.972
已知普通源	~10s	0.941	0.897
已知爆发源	~60s	0.955	0.941
未知暂现源	~90s	0.923	0.873

AI Agent 判断结果

判断结果

恒星耀发

判断标准:

候选体恒星耀发分析报告

根据EP观测的X射线暂现源候选体位置（RA = 304.969°, Dec = -9.607°），结合Simbad、AllWISE、Gaia DR3以及EP历史流量数据，对该源是否为恒星耀发进行综合分析。结论如下：

结论：该源极有可能是K型恒星的X射线耀发事件。

以下是详细的推理过程和数据证据。

1. Gaia DR3 数据分析：确认为一颗邻近的晚型（K型）恒星

Gaia数据在EP源位置附近（11.38角秒）探测到一颗恒星（Gaia DR3 6905471718711186944），其性质与恒星耀发的宿主星高度吻合。

Gaia DR3 参数	数值	分析与意义
角距离 (Angular Distance)	11.38 角秒	与EP源位置高度吻合，是可靠的光学对应体。
有效温度 (Teff)	4372.9 K	典型的K型主序星温度范围（3500-5000K）。
光谱型估计 (Spectral Type)	K-type	K型及M型矮星是已知最活跃的X射线耀发恒星。
视差 (Parallax)	2.371 ± 0.036 mas	视差信噪比高达65，测量非常精确。
距离 (Distance)	~422 pc	属于邻近恒星，符合"邻近"(nearby)标准（<1kpc）。
RUWE (天体测量单位权重误差)	1.045	<1.4，表明天体测量解是干净的单星，非双星或星系。
G, BP, RP 星等	G=15.61, BP=16.42, RP=14.73	颜色（BP-RP=1.69）与K型星一致。

2. X射线光度分析：光度水平与恒星耀发相符

利用Gaia测得的距离和EP观测的X射线流量，可以计算出该源的X射线光度。

X射线分析参数	数值	分析与意义
EP观测流量 (Flux)	3.79×10^{-11} erg/s/cm²	当前观测到的X射线流量。
推算X射线光度 (Lx)	~8.07×10^{32} erg/s	远低于典型活动星系核（AGN）的光度（通常 > 10^{35} erg/s），但完全处于K/M型恒星耀发的典型光度范围内（10^{28} - 10^{33} erg/s）。
历史流量对比 (Upperlimit)	历史数据不足	表明该源在以往的X射线巡天中未被探测到或处于宁静状态，符合"暂现"特征，是耀发事件的有力证据。

3. AllWISE 红外数据分析：排除AGN可能性

AllWISE在210角秒半径内探测到217个红外源，但无一符合AGN的红外颜色标准（W1-W2 > 0.8 且 W2-W3 > 2.2）。最近的几个红外源（如角距离11.3角秒、21.7角秒、24.0角秒处的源）的W1、W2星等均在15-16等左右，颜色正常，与K型恒星的红外辐射特征一致，而非具有热尘埃辐射的AGN。

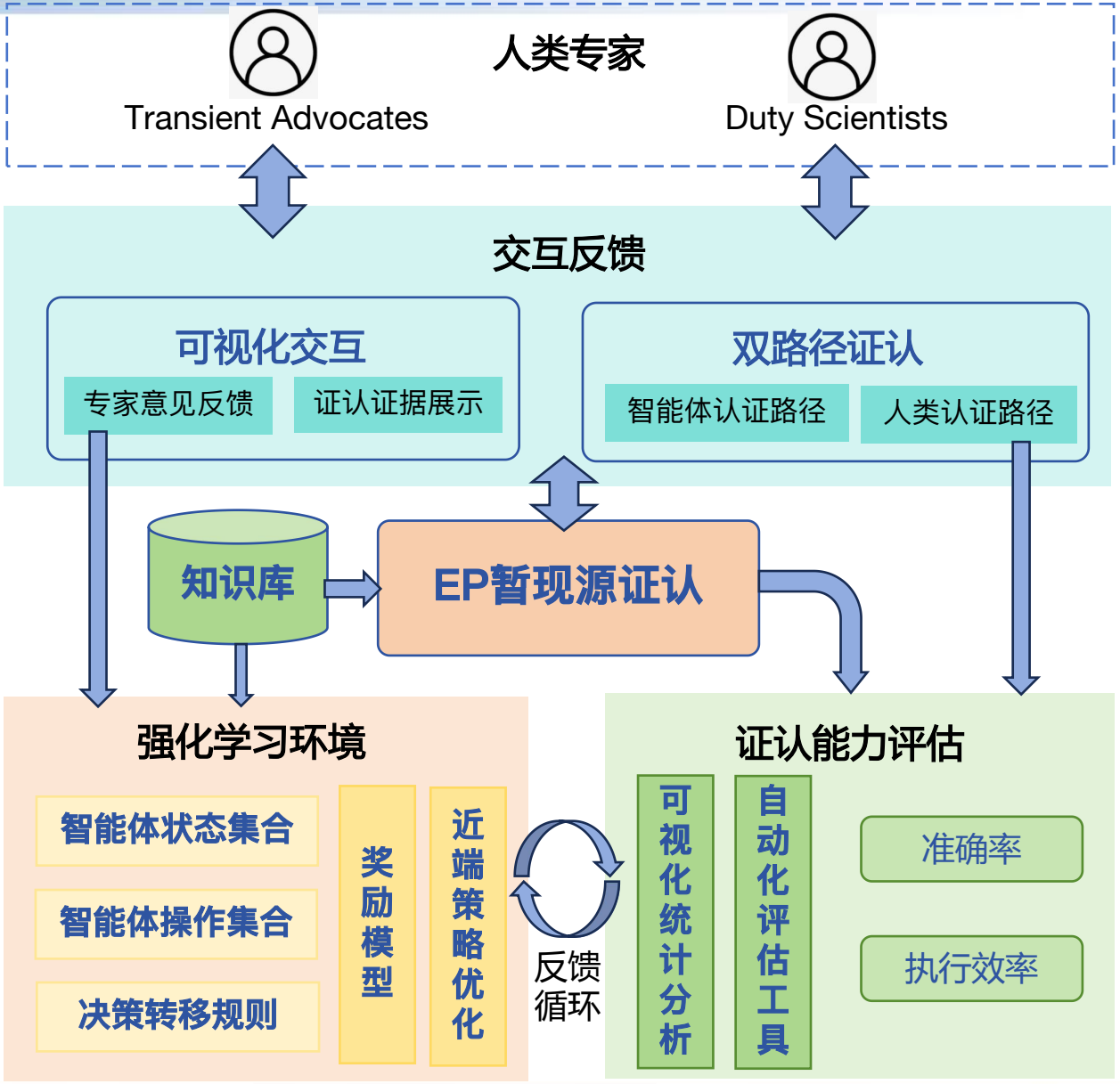
AllWISE 分析结果	数值/结论	分析与意义
AGN候选体数量	0	在搜索区域内未发现符合AGN红外颜色标准的源，基本排除了AGN的可能性。
最近红外源角距离	11.3 角秒	与Gaia源和EP源位置高度一致，很可能是同一个天体的红外对应体。

人在回路的智能体演进

National Astronomical Data Center
国家天文科学数据中心

- 强化学习
  - TA对结果给出打分并反馈
  - 训练奖励模型（7B）
  - 近端策略优化（PPO）更新底座参数
- 能力评估
  - 准确率、执行效率
  - ADT自动化评估工具





公开发布的NADC MCP Server

National Astronomical Data Center
国家天文科学数据中心

LAMOST数据获取工具：

根据obsid获取LAMOST低分光谱缩略图，中低分光谱的FITS文件、中低分光谱的JSON格式数据数组

AllWISE数据分析工具：

获取目标区域的AllWISE数据，通过红外颜色指数W1-W2和W2-W3判断目标区域是否存在AGN，用于判断是否为AGN变亮

Simbad/NED天体信息分析工具：

检索恒星、星系数据库，获取候选体误差范围内是否有已知天体，及其基本类型

高能历史观测流量分析工具

获取EP、Swift、XMM-Newton、Rosat观测记录，得到目标区域的历史流量上限值

Gaia恒星数据分析工具：

基于有效温度（teff）或bp-rb颜色指数判断是否存在晚型星（F, G, K, M型），判断是否为恒星耀发（距离小于1000pc， 光度小于1e35 erg/s）

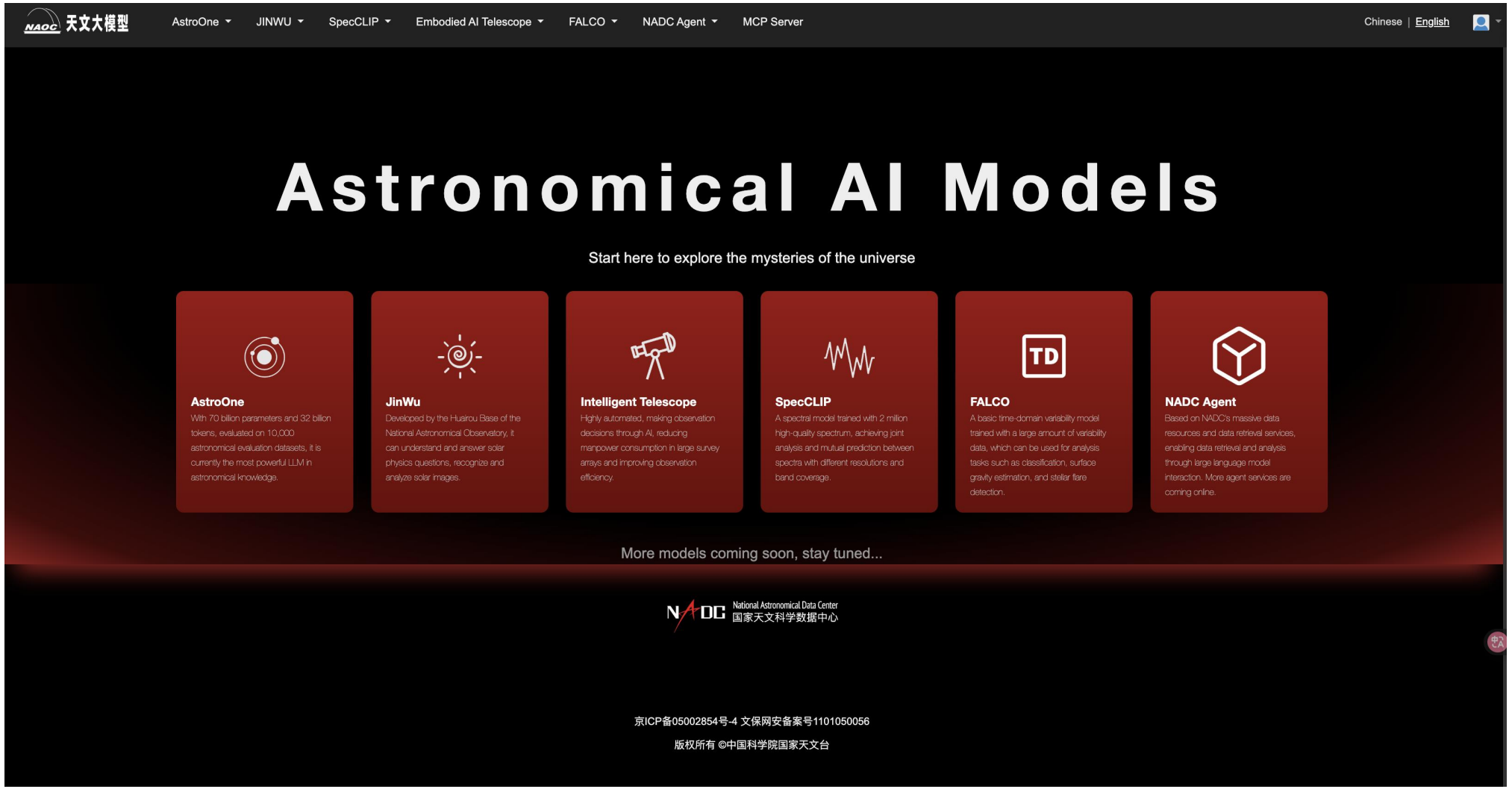
时域警报判断工具：

获取GCN、ATel、TNS在相关区域的时域警报，结合警报情况判断候选体类型

NADC AI大模型平台

National Astronomical Data Center
国家天文科学数据中心

注册平台账户，申请NADC MCP Server访问权限 https://nadc.china-vo.org/ai



总结与展望

National Astronomical Data Center
国家天文科学数据中心

- 基于NADC已有的数据融合服务，借助大模型的能力打造智能体群，用于提升时域天文的科学发现效率；
- 通过构建知识库，训练智能体，使其掌握时域天文知识，并知道何时应调用何种数据服务开展暂现源数据分析；
- 对NADC现有的数据服务进行改造，构建了基于MCP协议的数据工具，拓展了智能体的能力；
- 将进一步搜集科学用户需求，研发更多MCP工具和科研智能体，繁荣AI4S生态。

National Astronomical Data Center
国家天文科学数据中心

China-VO

谢谢！