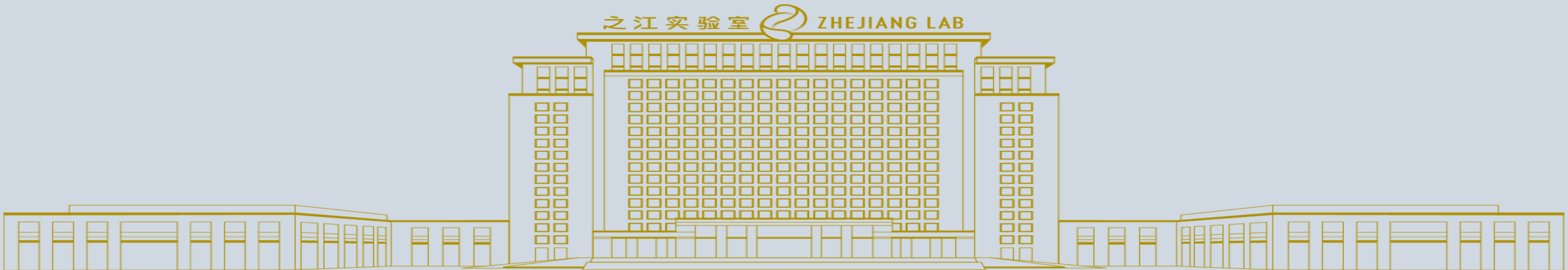
OneAstronomy: A Domain-Specific Large Language Model for Astronomical Science and Telescope Automation

Kuang Mao
Zhejiang Lab



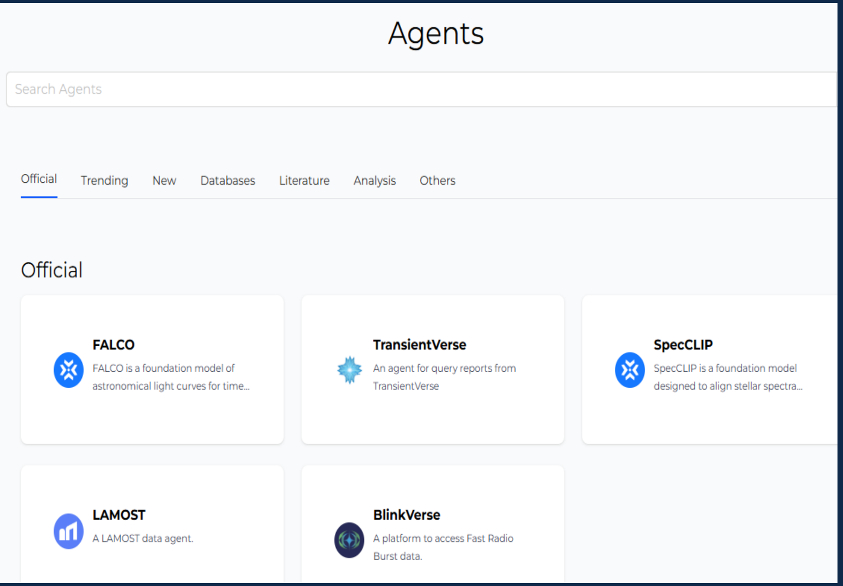
OneAstronomy’ s Components

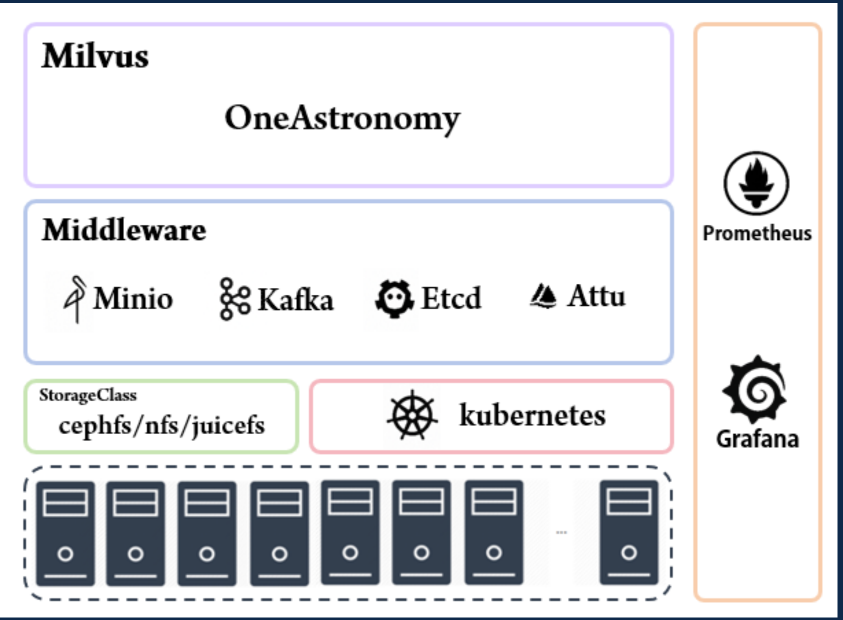
Agent System: Build on Openclaw’s Skill to manage complex workflows.

Domain-Specific Model: including an astronomical language model, SpecCLIP a spectroscopy model, and FALCO a time-domain photometry model.

Retrieval-Augmented Generation: We utilize a distributed Milvus Vector Database for high-performance retrieval.

Real-Time Access: The system is enhanced with web search for live astronomical data





Reasoning Model Training Process

Challenges in Astronomical Corpus Acquisition

The coverage of the astronomy domain and the quality of the corpus are the keys that determine the model's capabilities.

Basic Corpus- CPT corpus

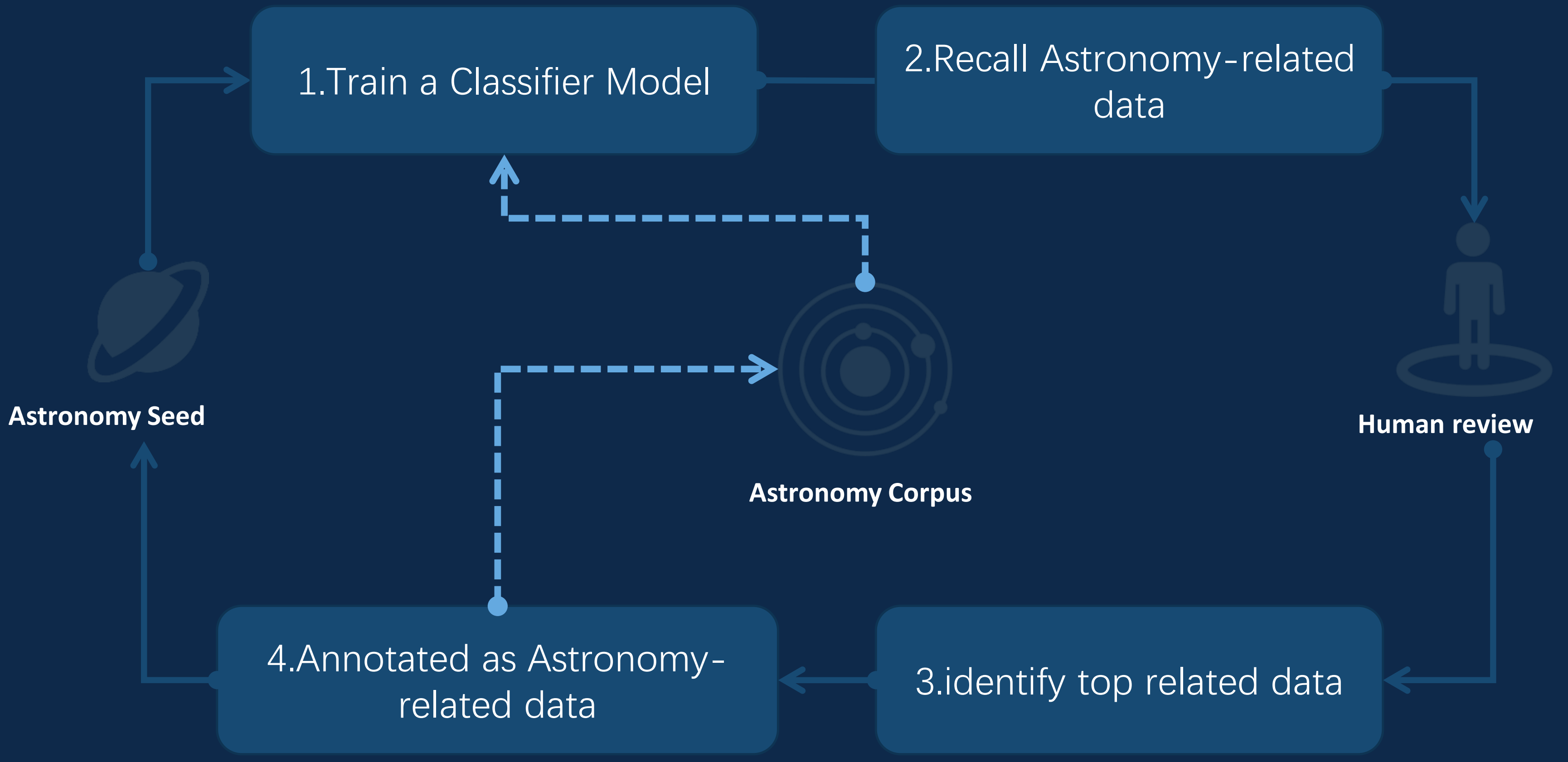
- domain-specific data is scarce and difficult to obtain;
- high-quality data is often unpublished;
- scientific research data is limited by copyright and open access agreements;

Astronomical reasoning QA Data – SFT corpus

- High-quality professional Q&A is scarce with mostly popular science content;
- Verifiable astronomy QA data are rare.

Basic Corpus Acquisition Pipeline

People from Zhejiang Lab and National Astronomical Observatory China participated in data collection and filtering, producing standardized Markdown-format data.



Corpus Parsing: Text, formulas, images, tables

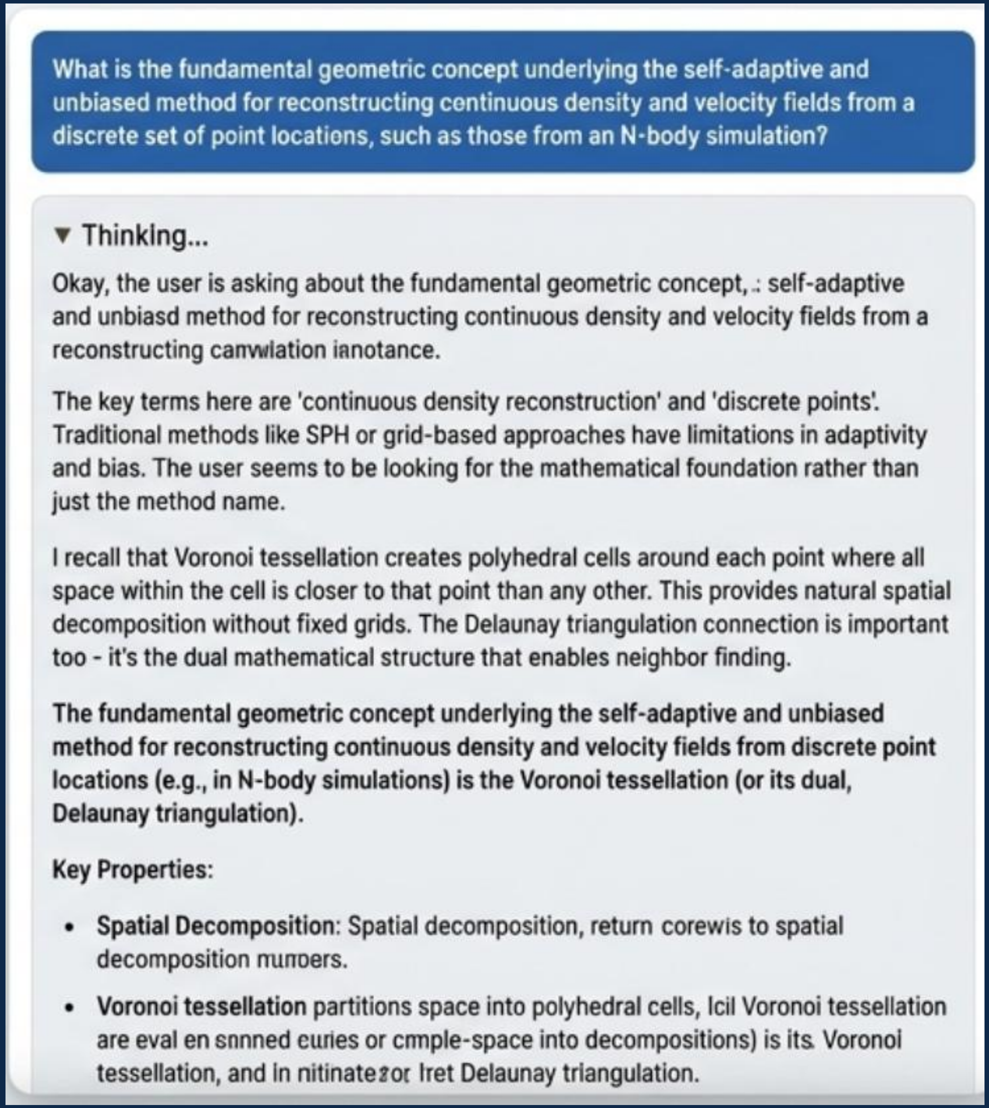
Data Cleaning: Rule-based & semantic filtering

Quality Assessment: Domain relevance, quality, language

Deduplication: Hash, similarity metrics etc.

The pipeline of data cleaning and processing

Reasoning Model Training: Q&A Pair Collection



OneAstronomy LLM is a reasoning model

Training Corpus Requirements:

- Accurate & professional verifiable Q&A pairs, multiple choice, true/false, calculation problems；
- High-quality reasoning chain-of-thought；
- High domain & knowledge coverage.

We have:

- Astronomical competition problems: IAO、USSAO、IOAA、BAAO；
- Open course exercises: university exam questions, astrophysics exam questions；
- QA pair in public benchmark.

Q&A Pair Synthesis Pipeline

Objective: Synthesize instruction pair data for SFT, while distilling reasoning processes to obtain QA pairs with Chain-of-Thought (CoT).

Step 1

Extraction

We extract high-quality, domain-specific Q&A pairs from the base corpus.

Step 2

Assessment

We evaluate and filter these pairs across multiple quality dimensions.

Step 3

Chain-of-Thought Extraction

We distill the LLM's internal reasoning process to serve as the thought chain for each question.

Step 4

Consistency Verification

We validate outputs and retain only questions with proven correct answers.

How we perform Q&A Distillation

- Classify our basic corpus by sub-discipline or paper type, such as observation paper, theoretical paper.
- Filter out all unprofessional content.
- Use different distillation prompts for different astronomical sub-disciplines

Distillation Prompt Template

```
You are an expert astrophysicist.
Your task is to read the following **Input text** from paper or book and generate quiz-style scientific Q&A pairs based on specialized astronomical knowledge summarized from the input text.

You must play **two roles**: the *Questioner* and the *Answerer*.
---
### Input text:
{paper_text}

---
### ROLE 1: Questioner
As the Questioner, carefully analyze the input text and generate **{question_count}} factual and specialized astronomy questions** that reflect the scientific content and findings, but do not explicitly refer to the input text itself.

Requirements for question generation:
({question_rule})
({common_question_rule})

Example of good questions:
({sample_question})

---
### ROLE 2: Answerer
Now act as the Answerer. For each of the above questions, write a precise and verifiable answer **based on the knowledge summarized from the input text**, ensuring that it reflects general scientific facts or reasoning rather than a specific citation.
citation.

Answering rules:
({answer_rule})
({common_answer_rule})

---
### FINAL OUTPUT FORMAT
Return your output as a single JSON object:

({output_format})
```

Thesis Classification	
Observational, Theoretical, Numeric[al,] Algorithmic, Review, Experimental	
Subject Classification	
Second-level Discipline English Name	T[illegible]
Celestial Mechanics	C[illegible]
Astrometry	A[illegible]
Cosmochemistry	C[illegible]
Galaxies and Cosmology	G[illegible] ar[illegible] O[illegible]
Stars and the Galaxy	Stellar Physics, Star Formation and Evolution, Physics of the Interstellar Medium, Structure and Kinematics of the Milky Way
The Sun and the Solar System	Solar System Formation and Evolution, Solar Physics, Planetary and Interplanetary Physics, Lunar Science, Meteoritics
Time Metrology	Time Metrology
History of Astronomy	History of Astronomy

Astronomy Paper Type

Q&A Pair Quality Evaluation

Quality metrics prompt: accuracy, completeness, relevance, clarity

1. Answerability (0–5 points)

* The question must be clearly stated, semantically complete, and unambiguous.
* Must not rely on external resources (papers, charts, appendices, data tables, etc.).
* Must be self-contained, including sufficient scientific background and context to be answered independently.
* If involving celestial objects, models, algorithms, or equipment, the specific object, parameters, or scope must be clearly defined.
* If containing formulas, they must be complete, standardized, use consistent notation, and include parameter definitions and units.

2. Professionalism (0–5 points)

* The question must fall within the professional scope of astronomy (e.g., celestial mechanics, stellar physics, galactic evolution, cosmology, planetary science, observational methods, etc.).
* Involves university or postgraduate-level professional content, not basic popular science.
* Demonstrates understanding of core astronomical theories, observations, or modeling research.

3. Expression Standardization (0–5 points)

* Must not contain vague references to an external source (e.g., "the paper", "this study", "the author", "in this section", "input text", "the figure", etc.).
* Professional terms must be spelled out, and undefined abbreviations are not allowed.
* Formulas must be expressed using standard LaTeX, with complete structure, unified notation, and clear parameter definitions.
* If the question involves multiple formulas, symbol consistency must be maintained, accompanied by parameter definitions.
* The question's grammar, logic, and format should conform to academic expression standards.

How to Evaluate

Manually Define Quality Filtering Rules
Initial quality criteria setup

↓

Improve Rules Based on Data Issues
Iterative refinement from existing problems

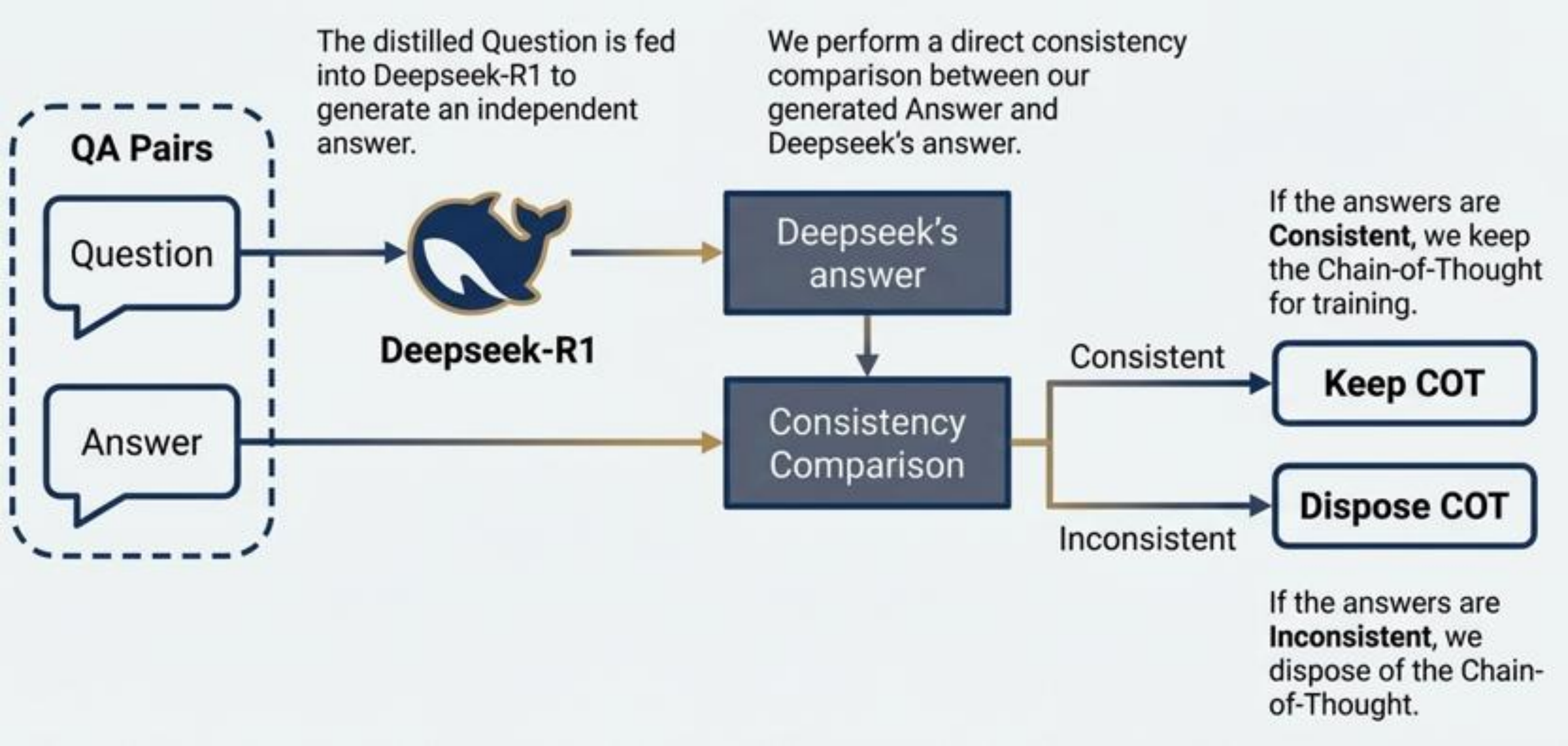
↓

Form Prompts for Large-Scale Evaluation
Automated assessment framework

↓

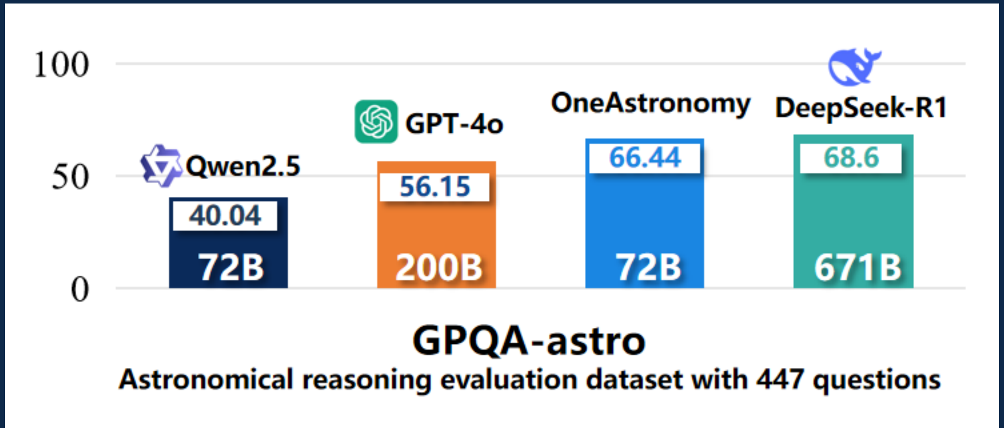
Use Large Models for Large-Scale Evaluation
Professional assessment output

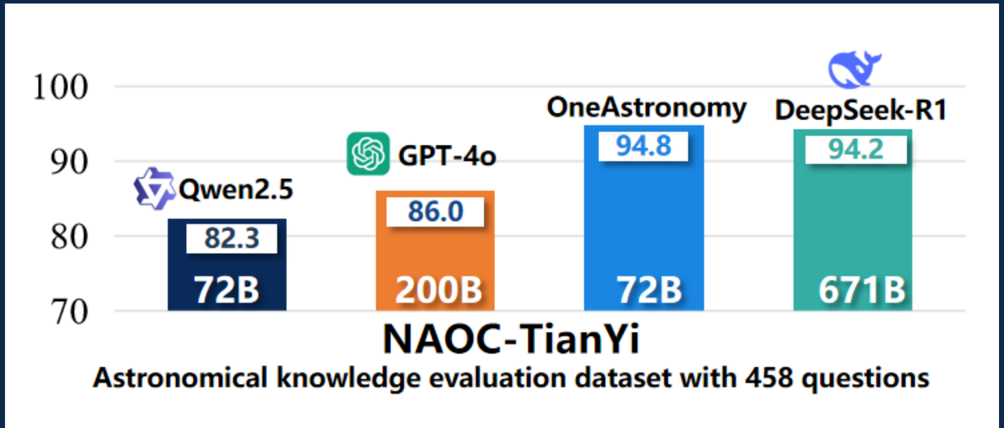
Thought-Chain Extraction and Consistency Checking

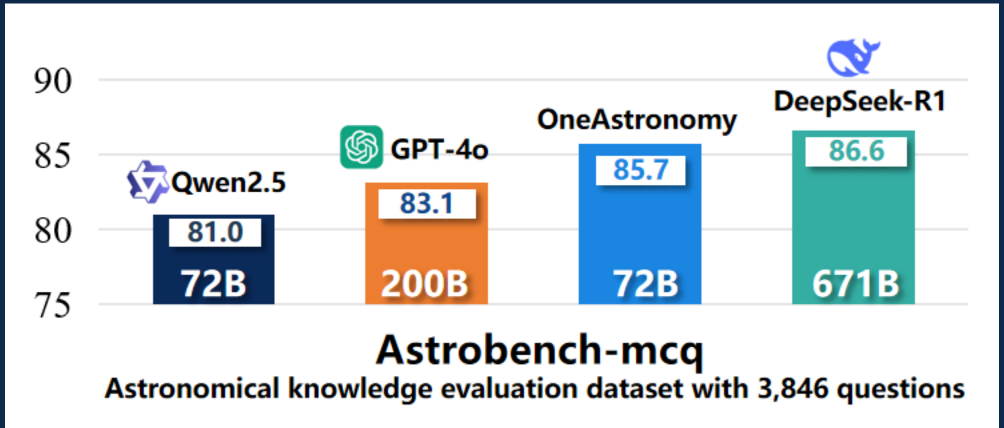


Performance Evaluation

We found that the OneAstronomy 72B-parameter model reaches the performance of large-parameter LLMs such as DeepSeek R1 (671B parameters)







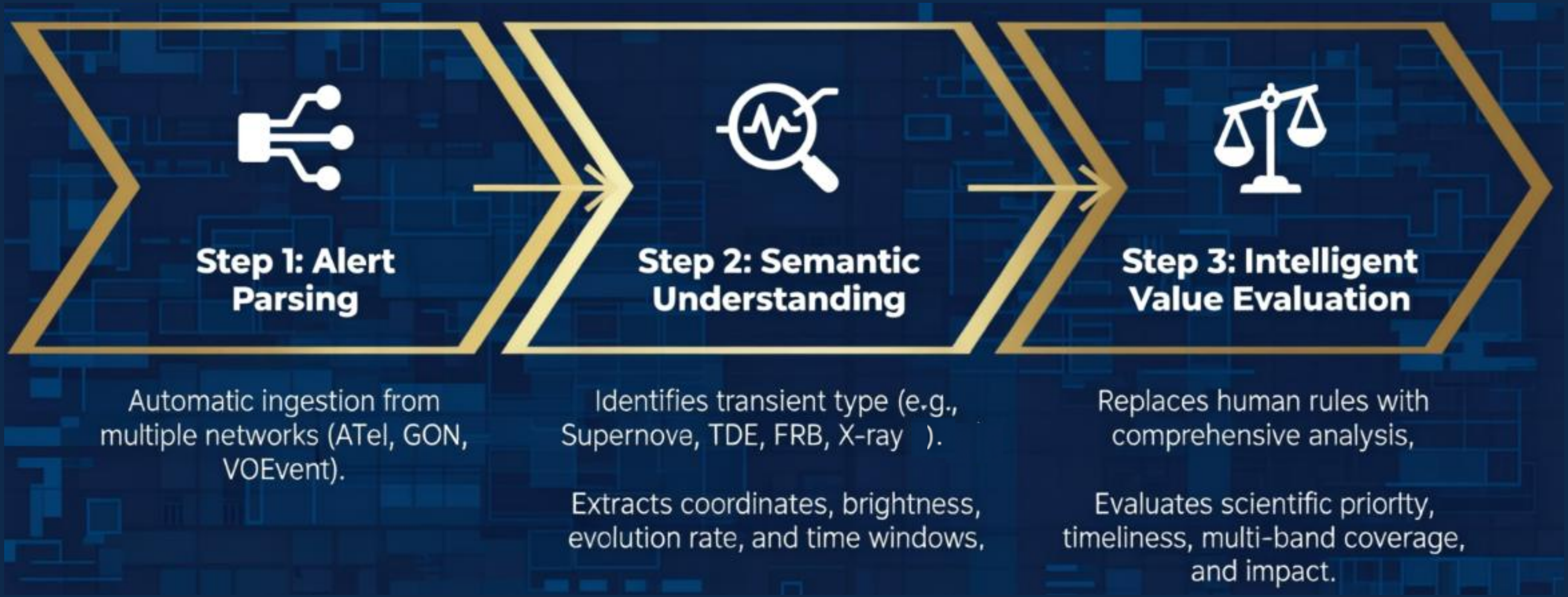
问题类型	OneAstronomy	OneAstronomy+Search
计算类	7	7.8
科研相关	7.4	7.7
实时科研相关	6	7.3
通识类	7.3	8
推理类	6.9	7.5
总体	7.1	7.7

Decision and Scheduling Ability for Transient Sources

Current Telescope Methods	Model-Based Embodied Intelligence
Observation Value: Manual experience-driven	**Observation Value:** Data-driven & intelligent reasoning
Scheduling: Non-real-time scheduling	**Scheduling:** Real-time dynamic scheduling
Control: Dependent on human operators	**Control:** 24/7 autonomous operation
Coordination: Single-station, local scheduling	**Coordination:** Multi-station collaborative scheduling

A fundamental shift from passive, manual scheduling to active, model-driven observation.

Model Capability: Analyze Observation Value



Model Capability: Observation Plan Generation

1. Physical Constraints Processing

- Accounts for astronomy, weather, and mechanics,
- Understands **each** telescope field-of-view and real-time status.

2. Survey Sequence Generation

- Translates scientific priority into exact sequences.
- Minimizes slew time and maximizes sky coverage depth.

3. Dynamic Rescheduling

- Instantly adjusts existing plans when high-priority transients trigger.

4. Multi-Site Coordination

- Treats the entire global network as one intelligent system.
- Balances load and enables cross-station relay observation.

PrestoZL: A GPU-Accelerated Jerk Search Toolkit for Binary Pulsars Search

- PrestoZL achieves end-to-end speedups 51.2 ~ 56.4 times over CPU-based Jerk Search in PRESTO with OpenMP;
- Signal discovery capabilities are fully consistent with PRESTO

Method	Process Time	Speedup
PRESTO C	9115m 22s	1x
PrestoZL	178m 3s	51.19x
PrestoZL-pipeline	161m 41s	56.38x

Figure 2: The time to process the 1816.77-second observation in M5(NGC 5904) and speedup relative to PRESTO C with OpenMP , jerk search parameters zmax=200,wmax=300

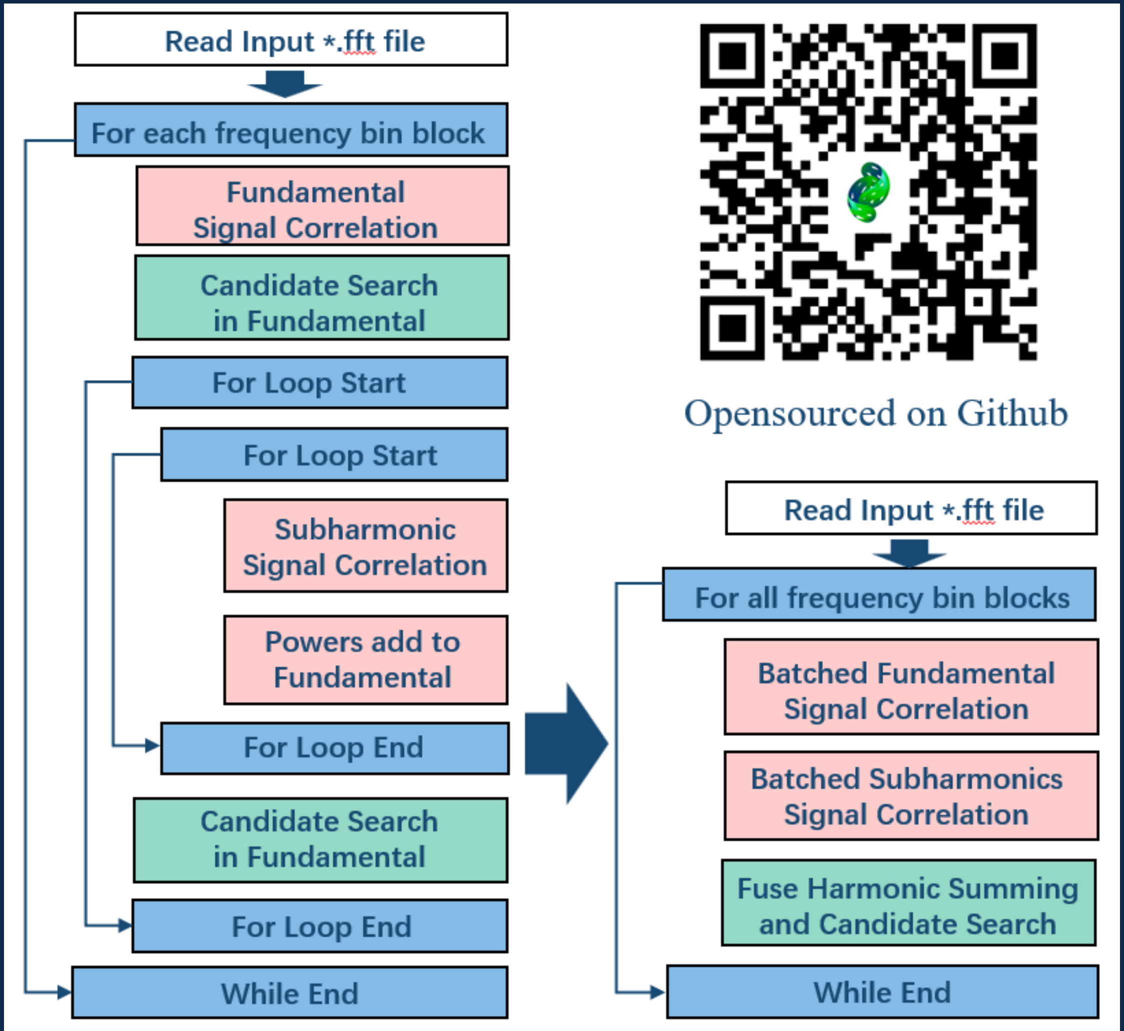


Figure 3: The computational workflow of CPU-based Jerk Search in PRESTO(left) and GPU-based Jerk Search in PrestoZL (right).

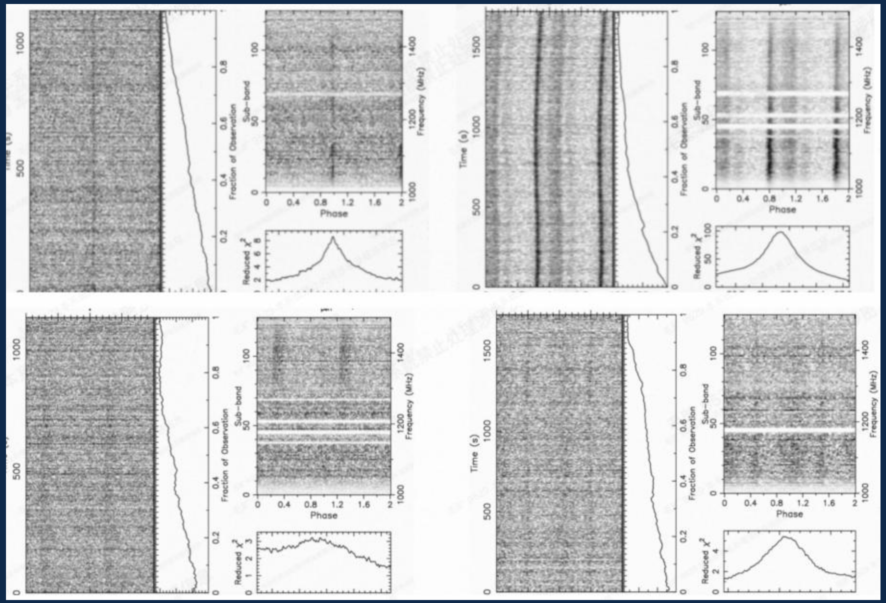
Kuang Mao, Scott M. Ransom, etc: PrestoZL: A GPU-accelerated High-throughput Jerk Search Toolkit for Binary Pulsars, APJS 2025

Pulsar Candidate Identification model: Pulsar-VLM

- We train a visual language model to identify pulsar candidate from PrestoZL's output
- Teach the model to think like an astronomer using language: vertical band structures, peak at a non-zero DM value
- Model was openweighted on huggingface: https://huggingface.co/jcchtt/Pulsar-VLM

Method	FN	FP	Recall	Specificity
PICS	29	4001	91.1%	68.1%
PICS-ResNet	51	4415	84.4%	64.7%
H-CCNN	4	2262	98.8%	81.7%
SGAN	31	2469	90.5%	79.2%
Pulsar-VLM	1	1884	**99.7%**	**85.2%**

Compared to traditional methods, Pulsar-VLM achieves higher recall and specificity.



Discovered four new millisecond pulsars from Fermi gamma-ray source

Kuang Mao, etc: Pulsar-VLM: A Visual Reasoning Model for Pulsar Candidate Identification, appear on APJS soon


Thanks