

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Agentic AI and Deep Learning at BESIII

- with a focus on potential new approaches

李科（like@ihep.ac.cn）

Institute of High Energy Physics

AI at lepton colliders

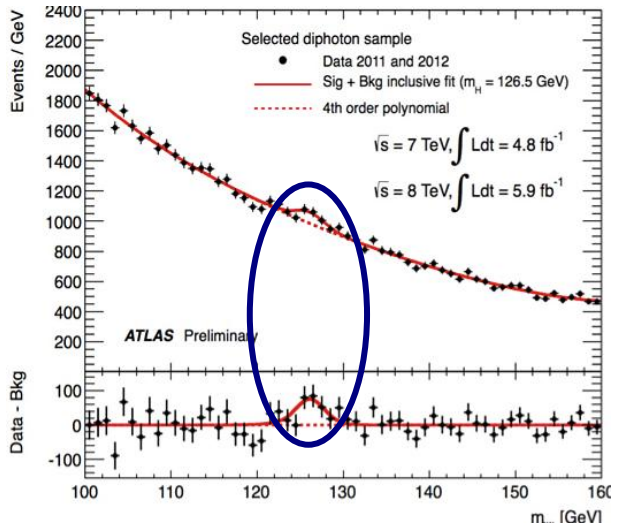
More than 10B events

Data analysis

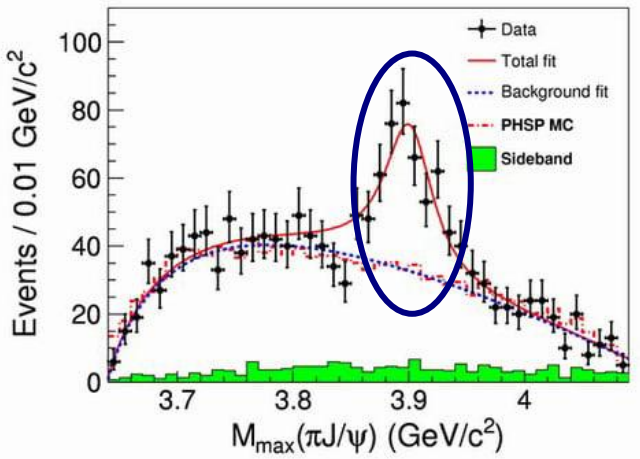
Significance $$Z \approx \frac{S}{\sqrt{B}}$$

AI4S 3 main approaches:

- **Improve S**
- Reduce B (more important at LHC)
- **Accelerate analysis**



Higgs at ATLAS



Zc(3900) at BESIII

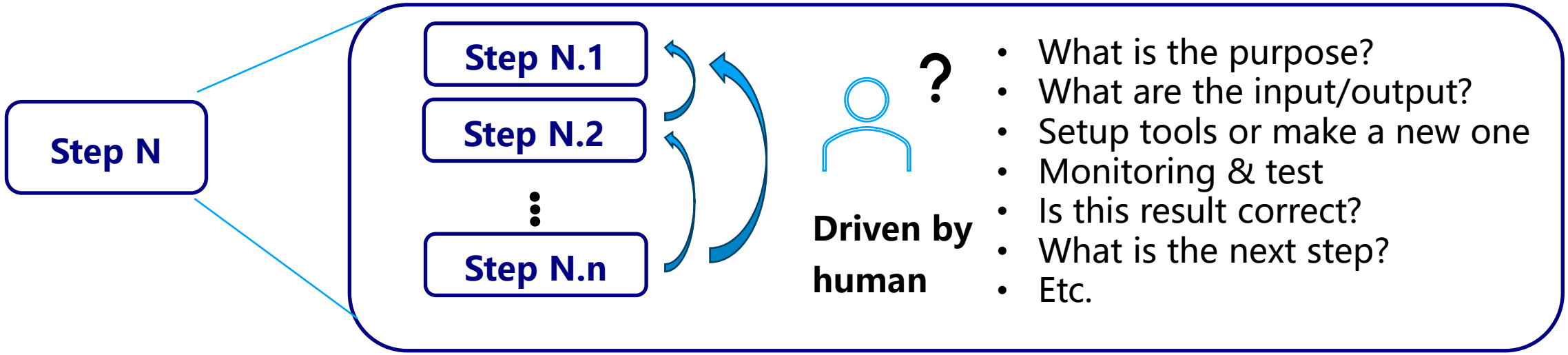
Outline

- New ways to search for new particles
  - Accelerate physics analysis: DrSai
    - The systematic search by automating the full analysis
  - Improve S: Foundation model and resonance tagger
    - The unified event representation model to cover all known physics channels
- Summary and outlook

Agentic AI for HEP

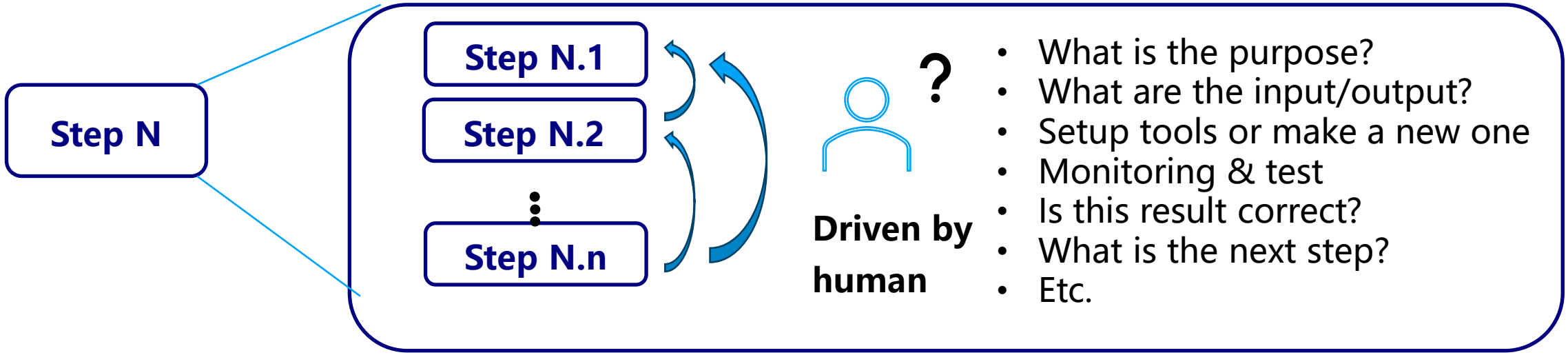
Traditional approach





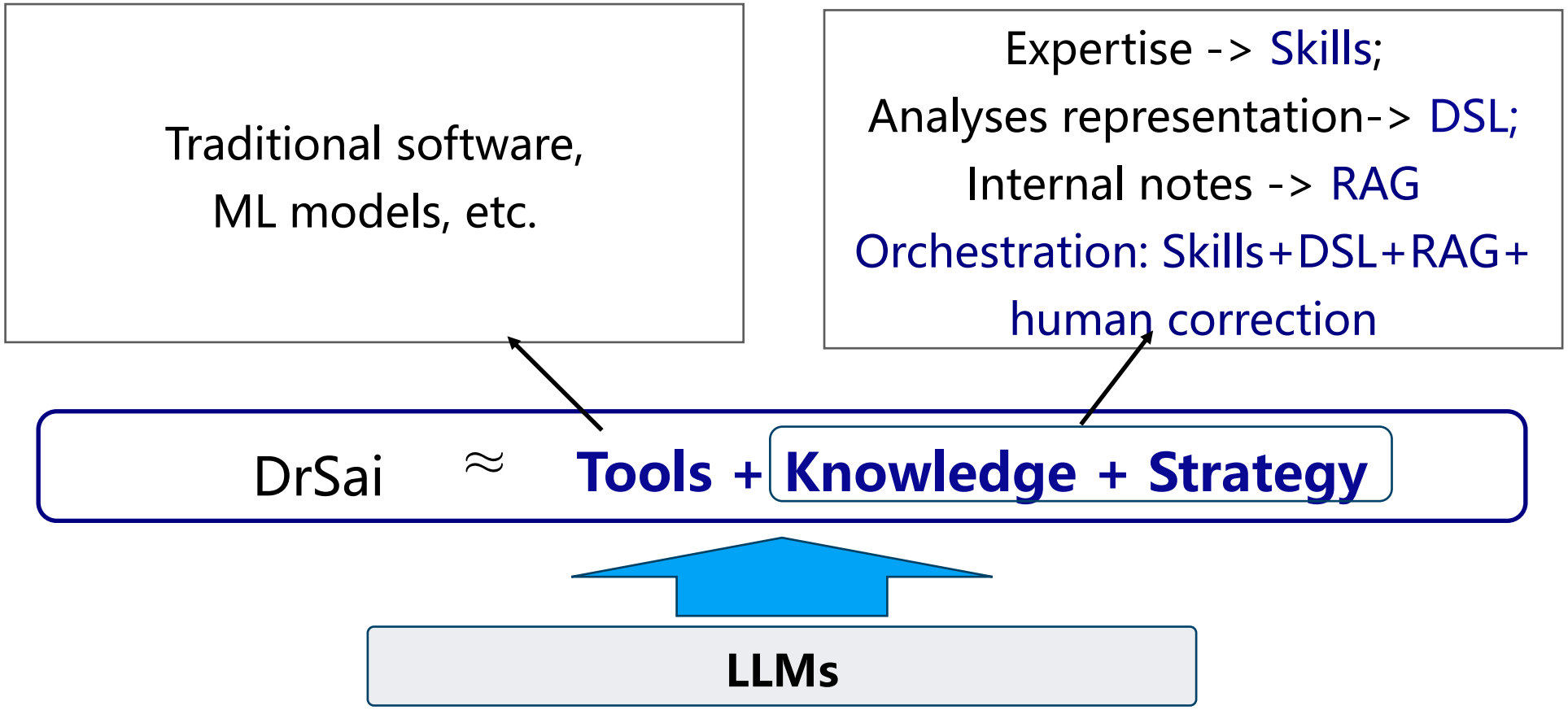
Agentic AI for HEP

- In traditional approaches, human involvement is required at all times.
  Expertise is often unstructured, though some can be represented by natural language.
- Today, we can leverage LLMs to build an AI partner for scientists.
  - **The AI partner designs and conducts the entire workflow based on human commands and feedback.**
- In the future, with a well-defined representation of knowledge, we can further develop an AI scientist.
  - **Humans would only need to propose a goal — and leave everything else to the AI.**

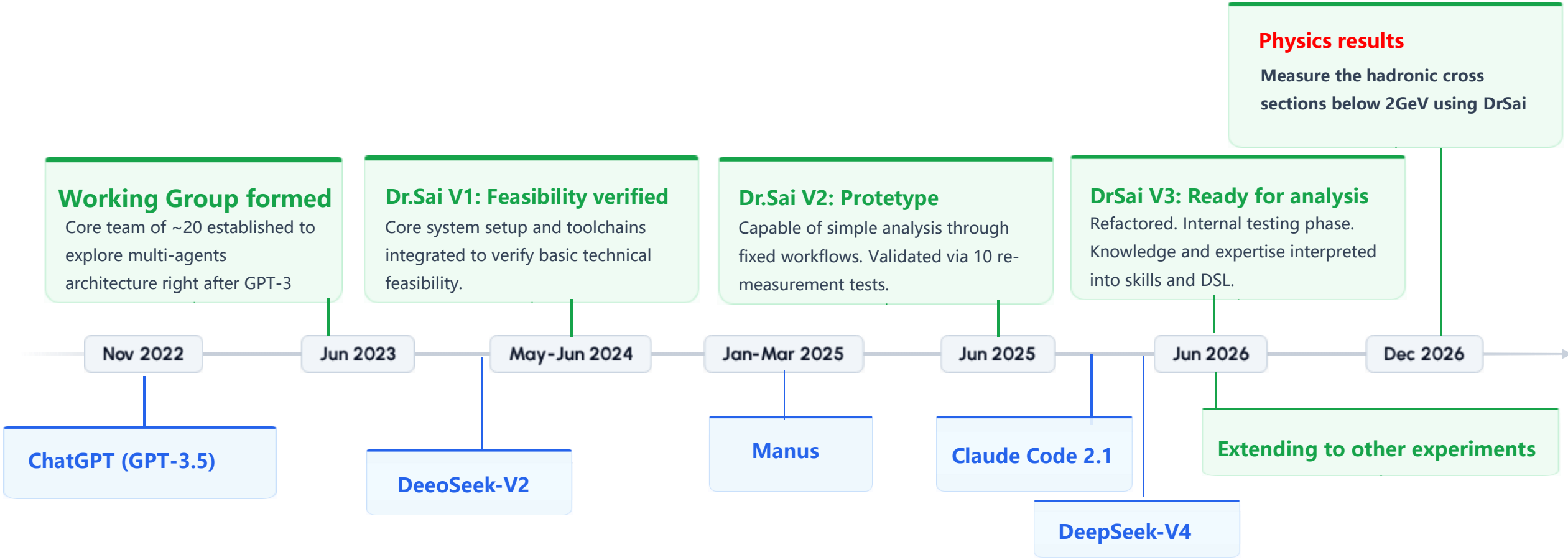


The DrSai project

Purpose: Pioneering autonomous physics discovery at BESIII (and beyond) through Agentic AI



DrSai Evolution Milestones



Identified Core Challenges

Knowledge & Self-Evolution

Often unstructured (experience, intuition). How to make it computable, reasonable, and continuously growing?

Solution

Domain-Specific-Language (DSL) represent analysis with an on-the-fly extraction and iterative refinement from expert feedback

Reasoning & Task planning

High level goals need complex and hallucinations-free decomposition + environment-aware reflections

Solution

Expert guides the system to orchestrate workflows for multiple analyses → distills into Skills → next workflow iteratively refined

Tools/codes & Execution

Complex software system, and must align with physics goals, as well as auto-checking

Solution

Error-Aware GraphRAG – stores long connections.

DSL/Skill align with code blocks or tools, enabling multiple checkpoints (e.g., injection tests).

Infrastructures & Safety

Scaling bottlenecks (GPU costs) vs. data privacy and autonomous safety risks

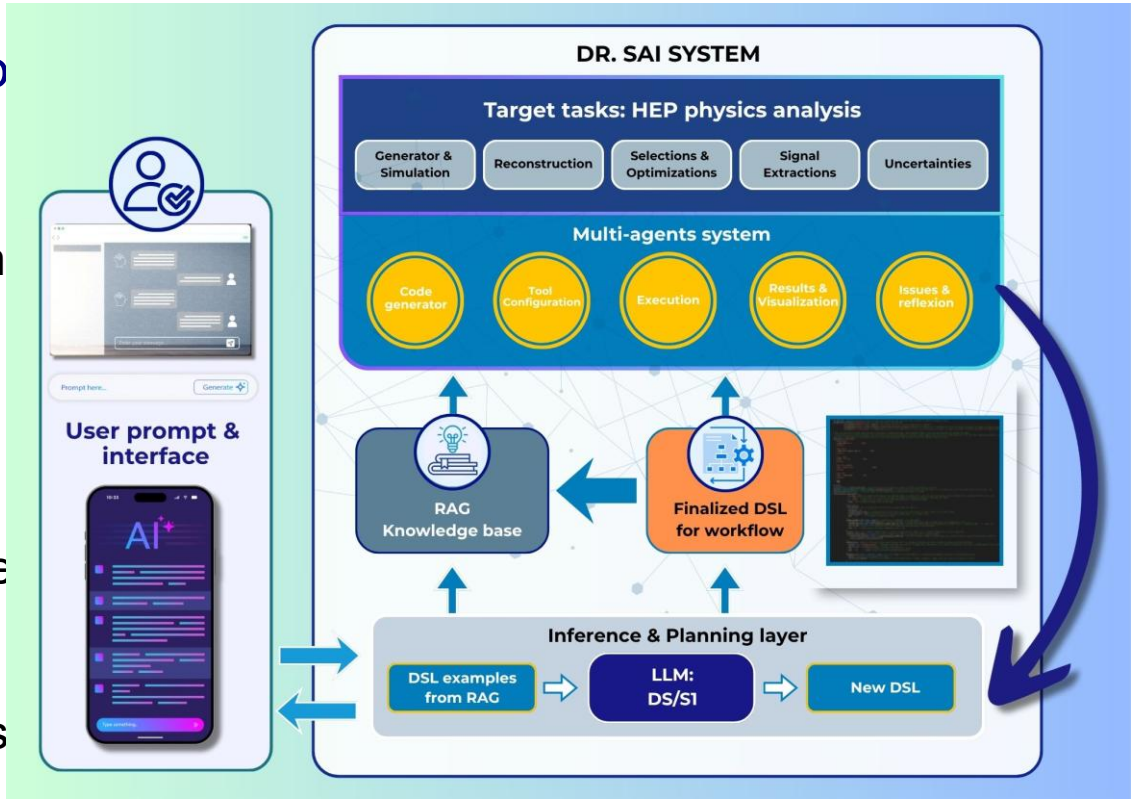
Solution

Local deployment of open-source LLM (DeepSeek) at IHEP + Cloud support.

Only allow predefined operations.

DrSai V3

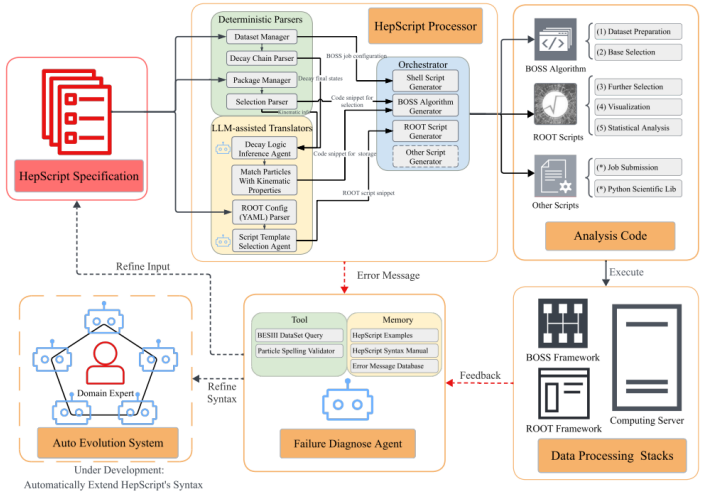
- With DrSai V2, we have identified four core challenges and developed corresponding solutio
- Major updates
  - AutoGen-based framework -> Claude Code agen runtime
    - Also DeepSeek-TUI, Cline, etc.
  - A set of analysis Skills – built and refined
  - Hundreds of DSLs extracted from published pape
  - Error-aware GraphRAG for code generation
  - Agentic RAG for internal notes of previous analys

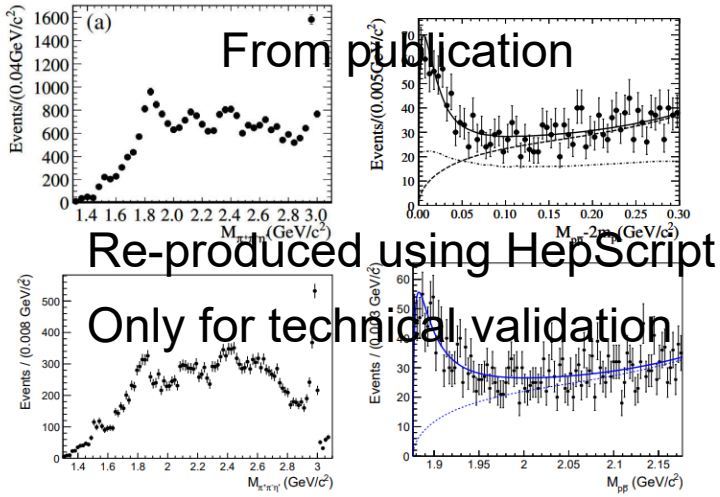


Front. Phys., 2026, 21(9): 096401

Analysis representation: HepScript (a DSL)

- Represent BESIII analysis
  - Pseudocode in Ruby/YAML format
- Bridges high-level intent with code/tools
- Defines a constrained action space
- The majority are mapped into code blocks
  - Only a small part need conversion by LLM
- **Scalable:** automatically extracted from published paper or generated on-the-fly for new analysis





HepScript	LLM	N_S	N_F	SR (%)	One Retry SR (%)	Three Retries SR (%)
LLM-generated (72 Packages)	Deepseek-R1	34	38	47.3 ± 5.8	87.8 ± 3.8	94.6 ± 2.6
	GLM-4.7	31	41	43.2 ± 5.7	90.5 ± 3.4	95.9 ± 2.3

Good accuracy for generation of analysis algorithms

arXiv:2605.01423

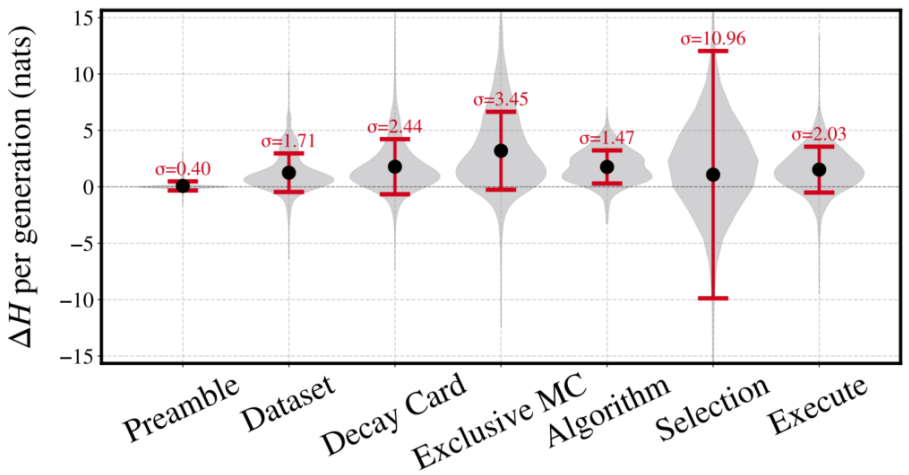
Knowledge base

- RAG + LLM-wiki, Content in knowledge base
  - Internal docs, e.g. memo and question-answer pairs from internal review
    - Chunked by analysis steps
  - Internal website and introdcutions
  - HepScript for published papers
- Introduce "systematic uncertainty" when building knowledge base
  - Define weights by "statistical" and "systematic" uncertainties

HepScript Example

Algorithm 1 Generic HepScript template.

Step 1 – Datasets
    $D_{\text{data}} \leftarrow$ DATASETMANAGER.REAL_DATA.FIND("703_4260") ▷ At 4.260 GeV
    $D_{\text{incMC}} \leftarrow$ DATASETMANAGER.INCLUSIVE_MC.FIND("703_4260")
Step 2 – Decay card $\mathcal{C}_{\text{sig}}$
    $\mathcal{C}_{\text{sig}} \leftarrow$ Decay $Y(4230)$: $\pi^+\pi^- J/\psi$ (PHSP) ▷ $e^+e^- \to \pi^+\pi^- J/\psi$
Step 3 – Exclusive MC D_{exMC}
    $D_{\text{exMC}} \leftarrow$ CREATE_EXCLUSIVE_MC(related $= D_{\text{data}}$, events $= 10^5$, $\mathcal{C}_{\text{sig}}$)
Step 4 – Algorithm config
    $\mathcal{A} \leftarrow$ ANALYSIS(name = "pipiJpsi", output = root)
Step 5 – Selection Pipeline ▷ Analysis-specific
    TRACK SELECTION
    PHOTON SELECTION
    PARTICLE IDENTIFICATION
    KINEMATIC FIT
Step 6 – Note (un-expressible cut)
    NOTES IN CONTEXT ▷ free-text physics constraints
Step 7 – Bind & Execute
  $\mathcal{A}$.BIND($D_{\text{data}}, D_{\text{incMC}}, D_{\text{exMC}}, \mathcal{C}_{\text{sig}}$)
  ROOT Files $\leftarrow$ $\mathcal{A}$.EXECUTE_ON($[D_{\text{data}}, D_{\text{incMC}}, D_{\text{exMC}}]$)
return ROOT Files

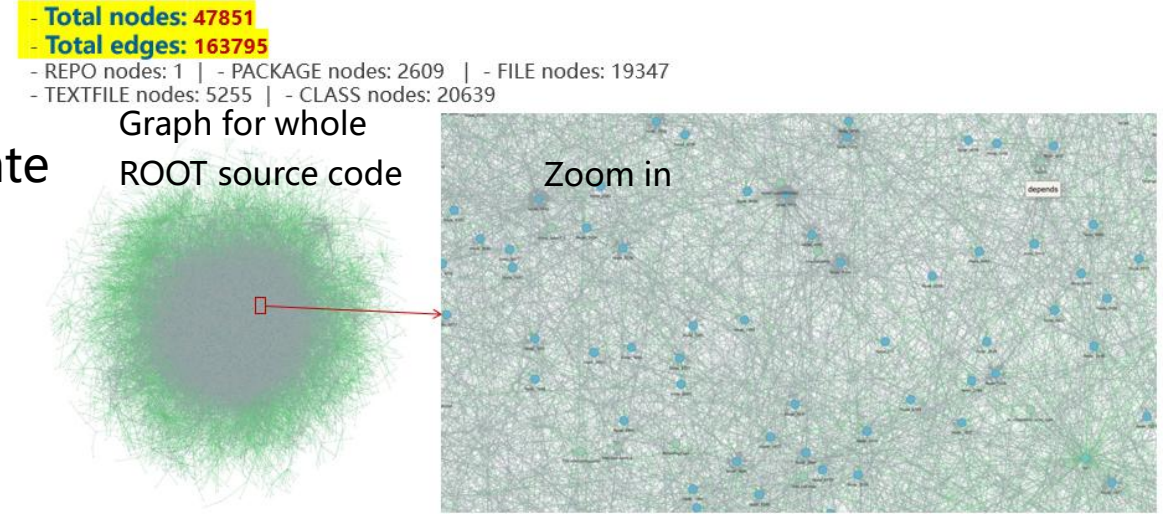


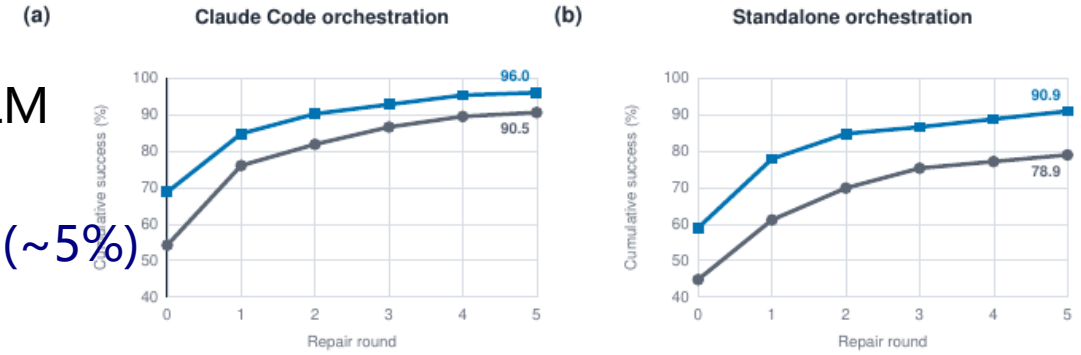
	Strict (%)			Struct (%)			Phys-Acc (%)		
Perturbation	NR	Rel	Conf	NR	Rel	Conf	NR	Rel	Conf
$k = 3$	71.8	76.2	**76.7**	89.3	91.0	**91.0**	57.6	62.4	**63.4**
$k = 10$	71.8	75.7	**76.3**	89.3	90.6	**90.8**	57.6	62.6	**63.3**
paraphrase	71.8	75.2	**76.0**	89.3	90.3	**90.7**	57.6	61.8	**63.9**
ORACLE		100			100			100	

arXiv:2605.01423

Domain code/script generation

- Complex code dependencies at HEP
  - LLMs struggle to reason over large-scale, intricate dependency correctly and efficiently
- We start with open-source software - ROOT
  - Massive interconnections -> a large graph
    - Nodes : class, package, etc.
    - Edges : contains, extends, depends
  - Obvious improvement on accuracy
  - Further improvement from Error-Aware RAG
    - First find the related subGraph, then feed LLM
- Success rate improved by ~10%
  - Still helpful for SOTA coding agent, claude code (~5%)
- Integrated to DrSai, works well
- This method can be easily applied to HEP software development



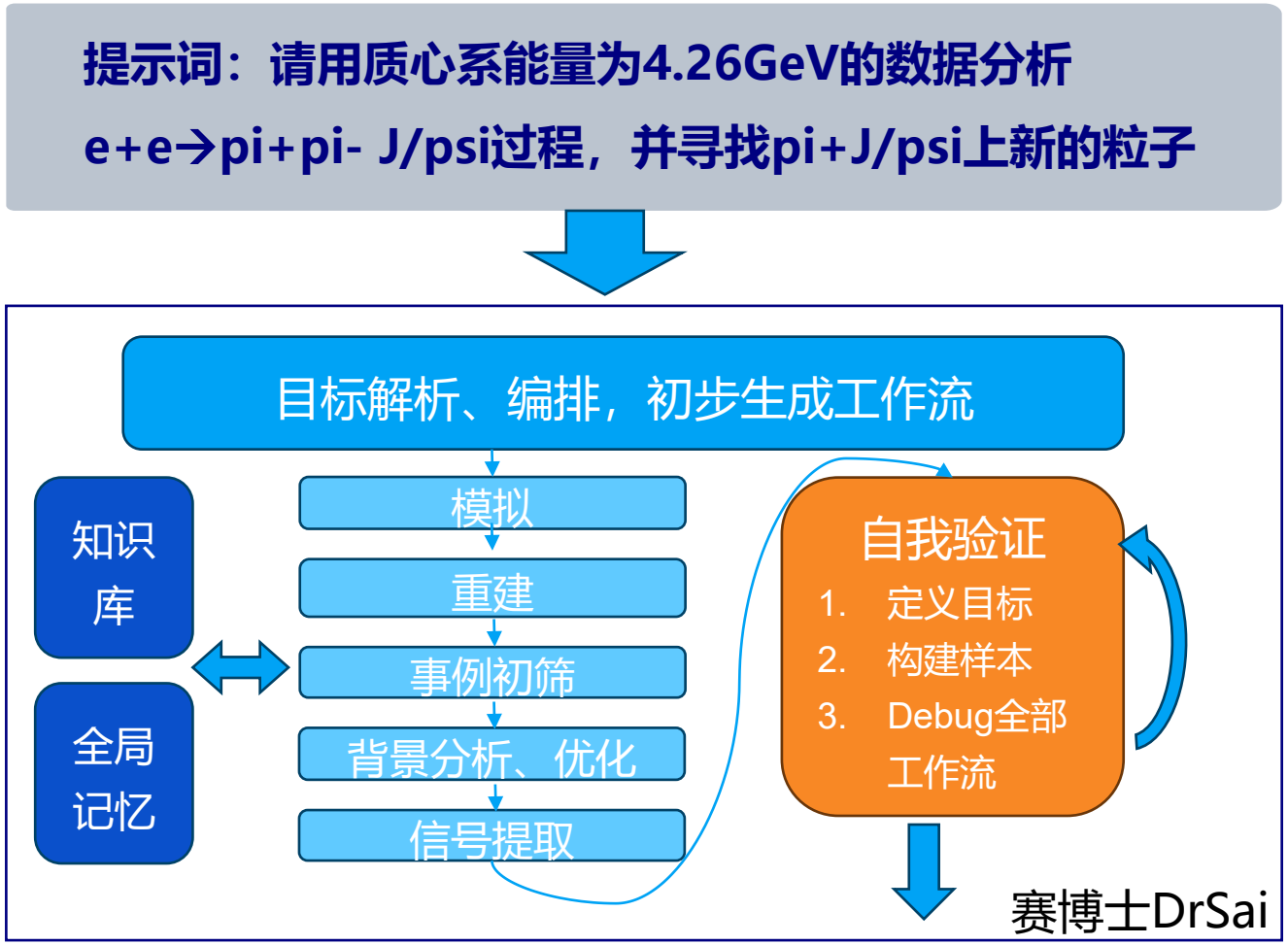


Github Repo: https://github.com/yues36067-cmd/ROOTER

arXiv:2609.01095

Status

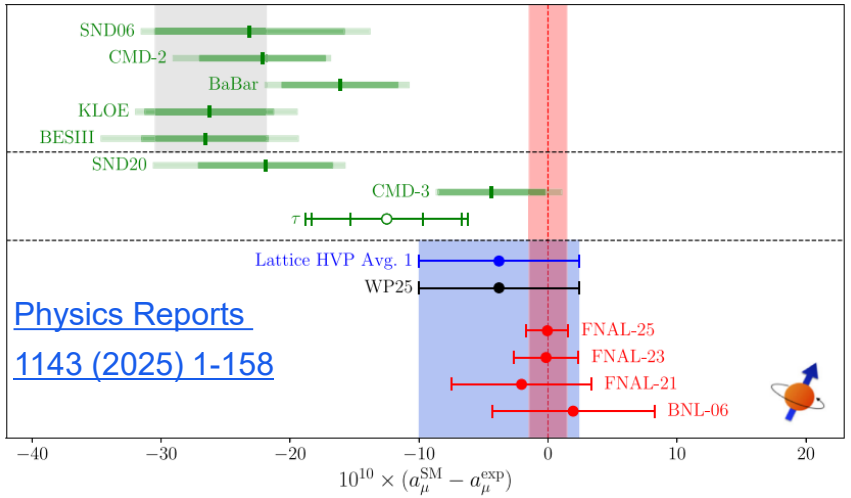
- Migrating the DrSai to DeepSeek Harness
  - Open-source, good maintenance
  - Consist physics results with Claude-Code
- Developing a new module: Truth-Anchored Recursive Self-Improvement
  - Use injection test to help the Agent to improve the whole workflow
- Many cutting-edge techniques
  - Welcome to collaborate

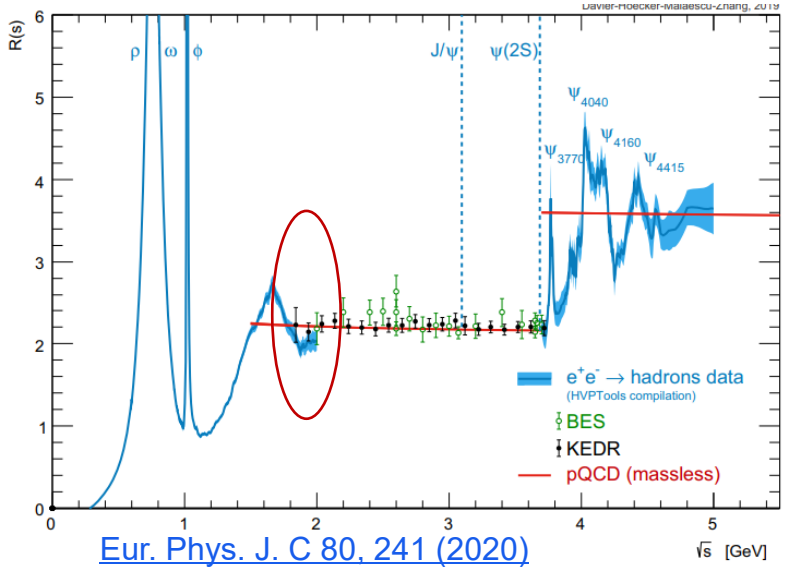


显著提升稳定性，目前可并行处理>30个物理过程

Analysis using DrSai at BESIII

- DrSai V3: **ready for BESIII physics analysis !**
  - **Vibe research – no coding needed now**
- **Now analyzing real data at BESIII:**
  - Measure all the hadronic cross section exclusively below 2 GeV, a painful analysis for human thus long delayed
    - A key input for HVP in muon g-2 calculation
      - Dispersive evaluation differ from LQCD
    - Experimental measurements of e^+e^- →hadrons are particularly important.
  - **Starting from 1.9 GeV:**
    - Experts steer the setup and input essential parameters, e.g. generators
    - DrSai design, execute and verify the entire analysis under expert supervision
    - Experts will **cross-check** all the measurements





AI at lepton colliders

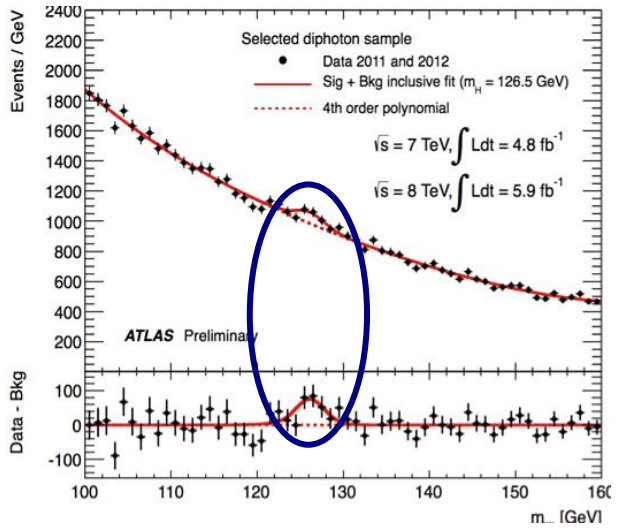
More than 10B events

Data analysis

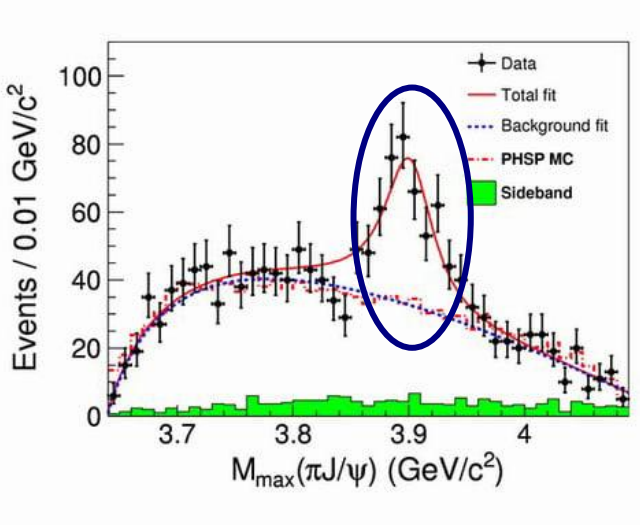
Significance $$Z \approx \frac{S}{\sqrt{B}}$$

AI4S 3 main approaches:

- **Improve S**
- Reduce B (more important at LHC)
- Accelerate analysis



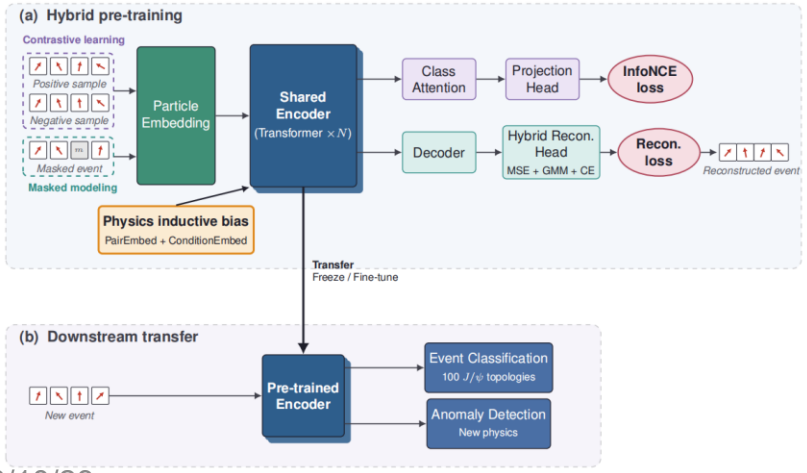
Higgs at ATLAS

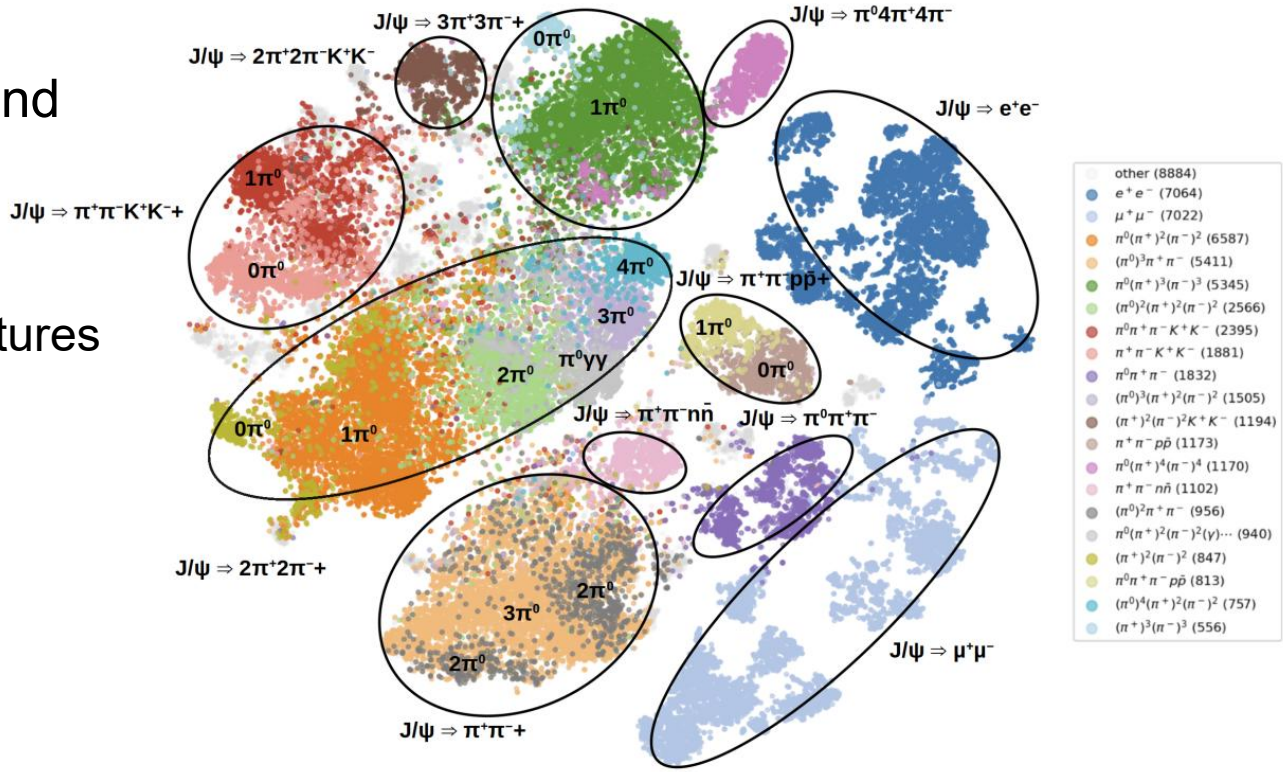


Zc(3900) at BESIII

Foundation model to classify all the events

- Goal: classify all the channels and support all downstream analyses
- Event representation using self-supervised and Contrastive Learning
  - Mask one particle then predict
  - Gate attention conditioned on event level features
  - Supervised fine-tune with a small dataset



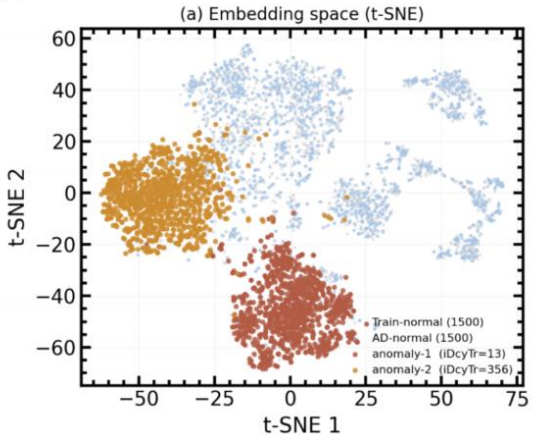


Foundation Model

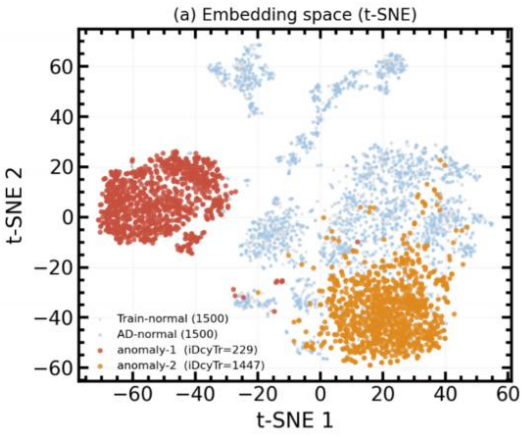
- Compared to cut-based analysis
  - Higher signal efficiency, cover all the decay channels
  - Lower background, all the bkg can be classified as well
  - Support almost all downstream analysis tasks
- Compared to ML-based analysis
  - No need to retrain the model
- For the new particle or channel
  - Auto-clusterization as anomaly
  - Different new particles or channels can be clusterized to different “shower”
- Now only use 10M events for training, will expand to 1B

Final state	Method	Precision	Recall	F_1
$\pi^+\pi^- p\bar{p}$	Cut1	6.3%	56.5%	11.3%
	Cut2	99.5%	50.9%	67.3%
	CLEAR ($\tau = 0.5$)	82.9%	92.5%	**87.4%**
	CLEAR (matched)	—	—	—
$\pi^+\pi^- K^+K^-$	Cut1	31.9%	55.8%	40.6%
	Cut2	99.3%	46.0%	62.9%
	CLEAR ($\tau = 0.5$)	92.8%	78.0%	**84.7%**
	CLEAR (matched)	99.3%	59.8%	74.6%
$\pi^+\pi^-\pi^0$	Cut1	79.0%	51.8%	62.6%
	Cut2	97.7%	37.2%	53.9%
	CLEAR ($\tau = 0.5$)	90.3%	90.9%	**90.6%**
	CLEAR (matched)	97.7%	46.0%	62.6%

(1). Near-OOD Detection

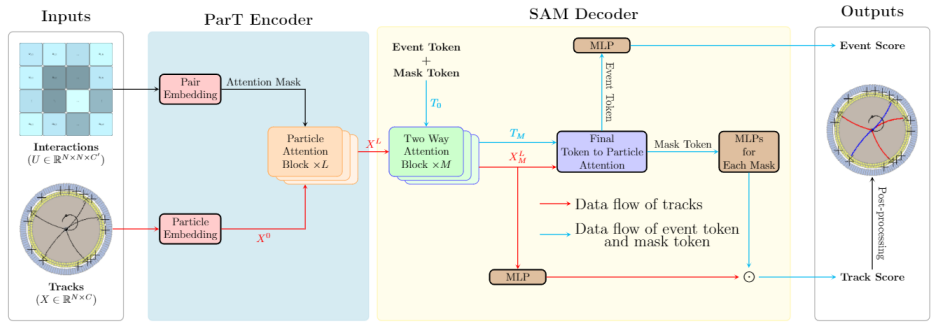


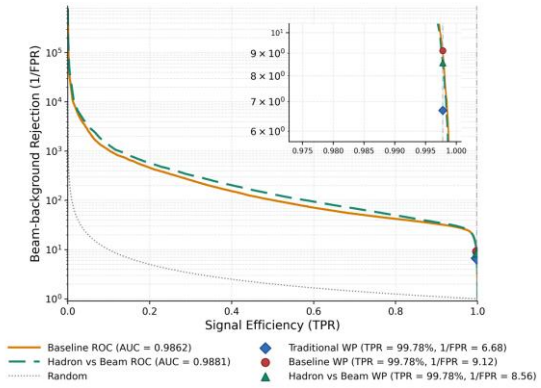
(2). Far-OOD Detection

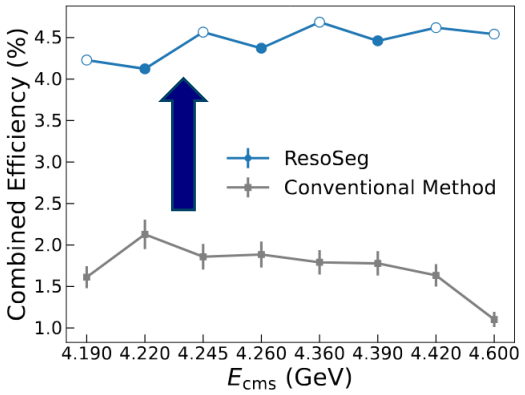


Other new results at BESIII

- Light-NN at FPGA to suppress the beam-induced background for trigger
- Trained on real experimental data
- **Background rejection improved by 20-50%**
- Latency is about 35ns
- Important for BEPCII-UP or next high-lumi lepton collider







Resonance Tagger

- Trained a latent space by Transformer
- Reconstruct the resonance as segmentation on final states (SAM model)
- **Statistics improved by 2.6 times**
- Migrating to D/Ds-tag

Summary and outlook

- Possible two new approaches for AI4S studies at lepton collider
  - Unique scenario w.r.t. hadron collider
  - Systematically analyze all possible process automatically
    - Autonomous analysis using DrSai - analyzing real data now
  - Use experimental data to build a large model - finalizing prototype
    - Classify all the events with foundation model
    - Tag the intermediate resonance by segmentation
- Welcome to join ! We are forming a DrSai collaboration now
  - Extending to: CEPC/JUNO/LHAASO analysis, etc.

backup