



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

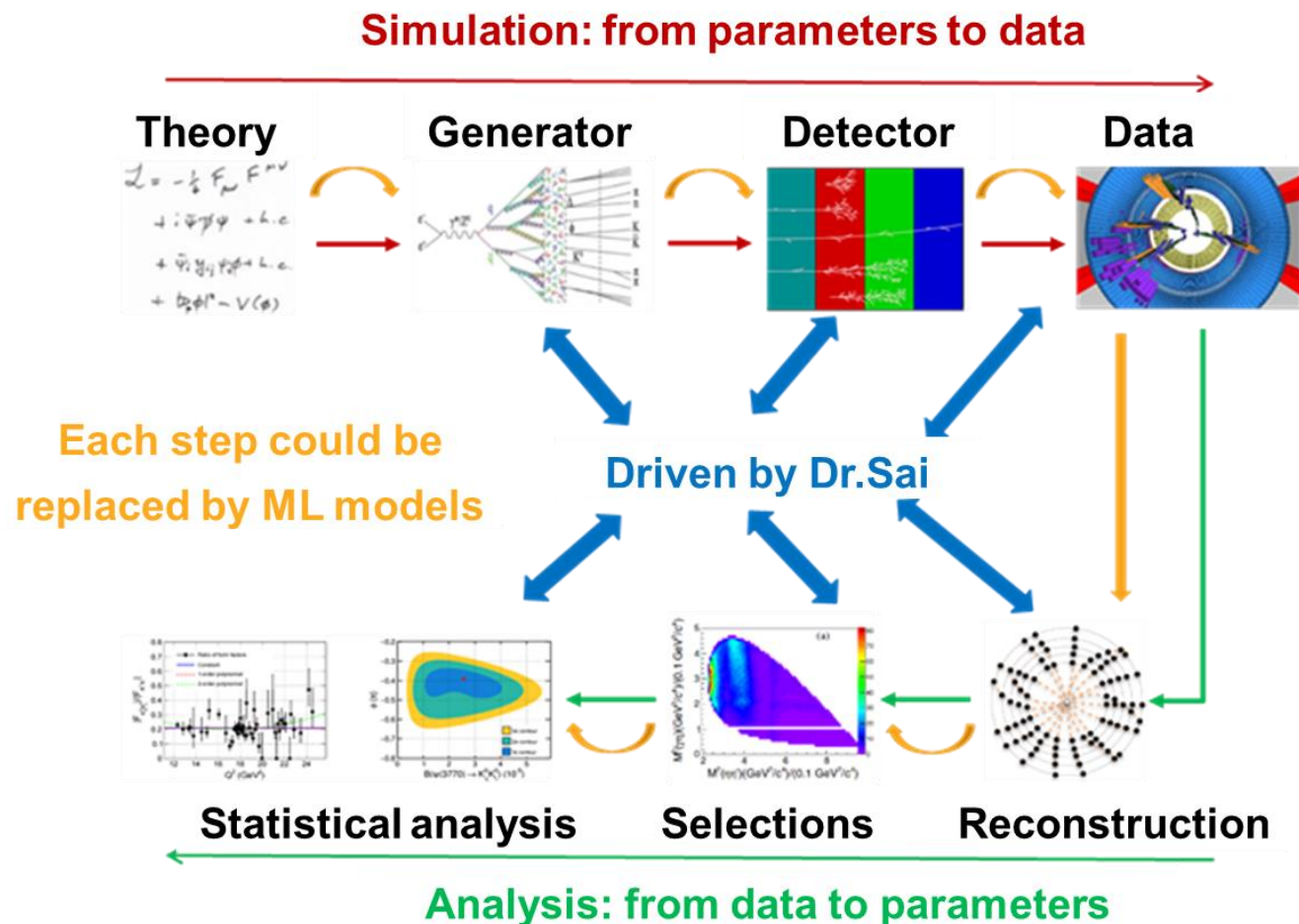
Leveraging Agentic AI for Global and Systematic New Physics Search

李科 (like@ihep.ac.cn)

Institute of High Energy Physics

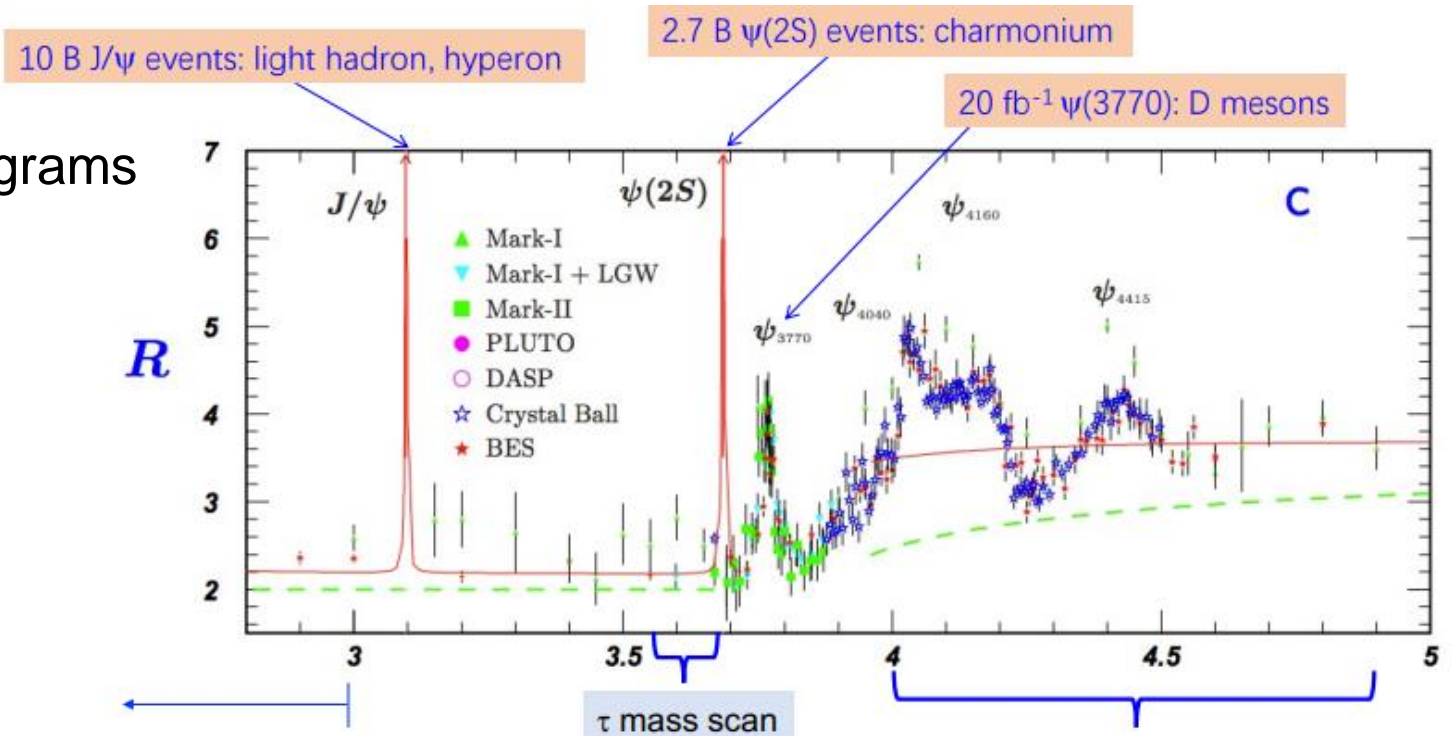
Outline

- New ways to search for NP
 - The systematic search with Dr.Sai
 - The global search with cross-dataset anomaly detection
- Summary and outlook



Introduction of BESIII - physics program

- >700 scientists and engineers
- Tau-charm factory, rich physics programs
 - Light hadrons
 - Charm meson/baryons
 - Charmonium
 - Precise test of SM
 - Search for new physics
- Hundreds of physics results



Low energy scan: R-value, baryon pair

High energy scan: exotic hadrons, charm meson/baryons

- Discovered >30 new hadrons
 - First tetraquark: $Z_c(3900)$

- **Clear and relative simpler workflow for effective modelling**
- **Different from LHC, unique challenges and advantages in lepton collider experiments**

Dr.Sai at HEP

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

Dr.Sai \approx **LLM + Tools + Knowledge + Strategy**

DeepSeek/
Qwen/GPT/
Gemini, etc.

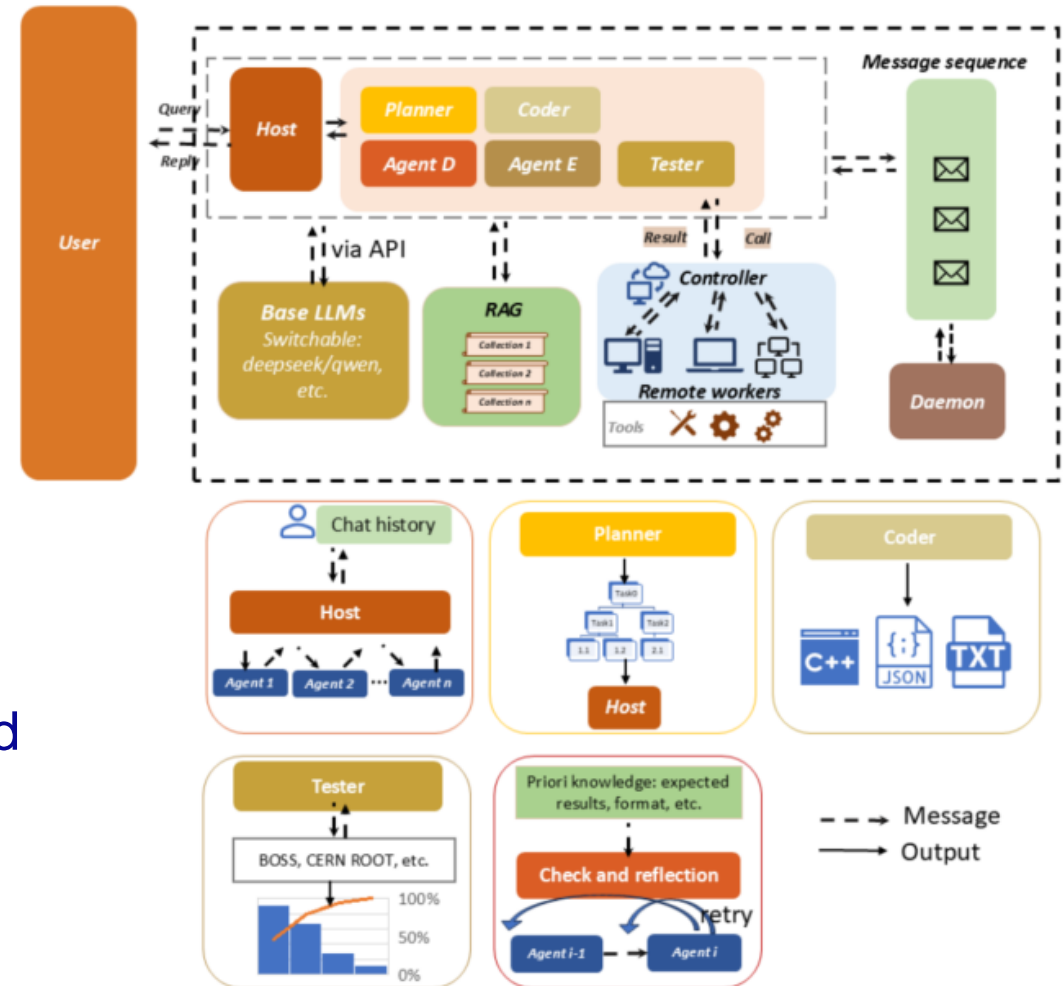
HPC, BESIII Offline
Software System,
HTCondor, CERN
ROOT, etc.

Previous analyses,
Skills + Domain-
Specific-
Language(**DSL**) +
RAG

Autonomous for
simple cases;
Collaborative for
complex ones

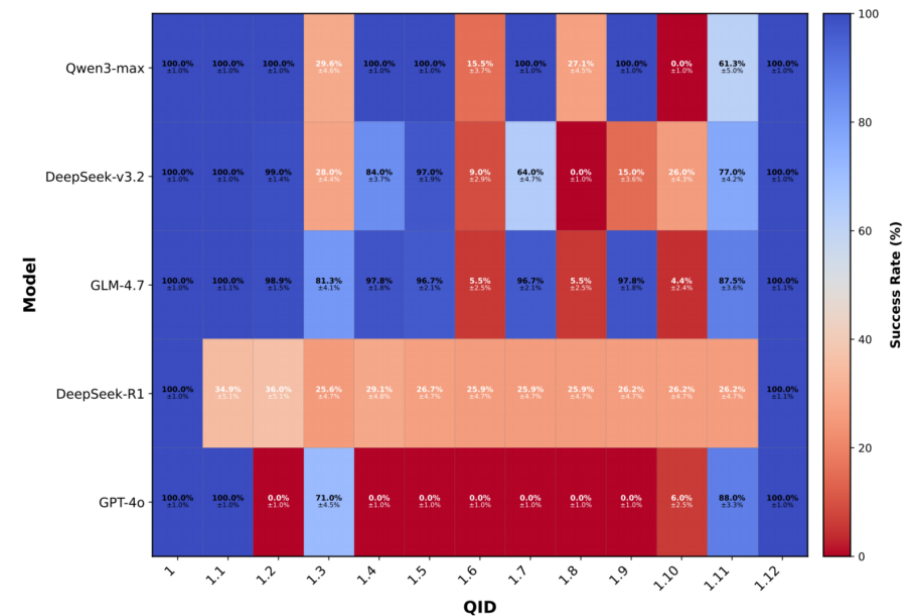
Dr.Sai at 2025.07

- A prototype, developed based on AutoGen framework
 - Collaboration with Computing Center
- Integrated to BESIII Offline Software System, IHEP HPC, and HTCondor system
- Re-organized analysis algorithms
- Pre-defined analysis workflow
 - Capable of **very simple analysis with fixed workflow**
- Base LLM is switchable for agents



Demonstration of Dr.Sai: re-measurements

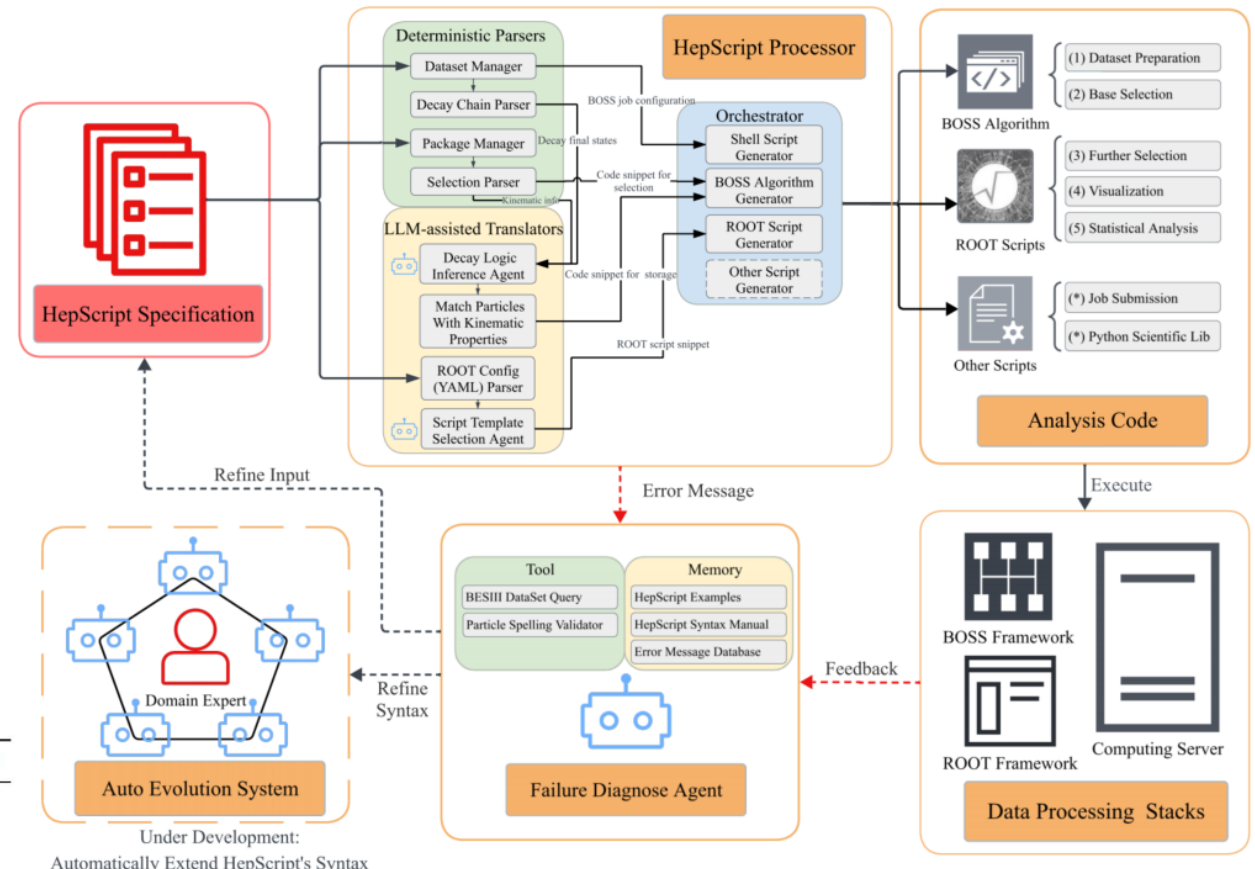
- We validated the whole system by re-measure branching fractions of J/psi decays !
 - Build an inclusive MC sample with all known J/psi decays + new physics signals
 - Use $\psi(3686) \rightarrow \pi^+ \pi^-$ J/psi process to measure the BFs, good consistency with injected BFs



DSL (HepScript) for BESIII analysis

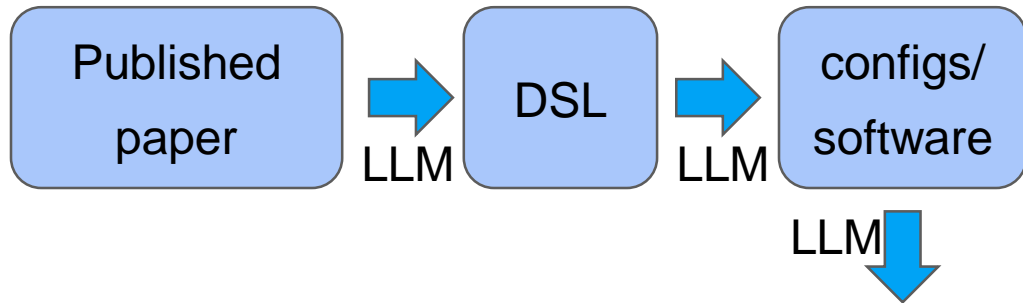
- A guide to agents/LLMs
 - Formalizing domain knowledge
 - Integrating with domain software/tools
 - Constraining action space
- Scalable and Automatically
 - Define preliminary DSL by human experts
 - Extract from published paper by LLM
 - Convert to codes/configs by LLM

HepScript	LLM	N_S	N_F	SR (%)	One Retry SR (%)	Three Retries SR (%)
Human-written (63 Packages)	deepseek-v3.2	62	1	96.5 ± 2.1	100 ^{+0.0} _{-1.5}	-
	deepseek-r1	56	7	87.7 ± 4.0	100 ^{+0.0} _{-1.5}	-
	gpt-4o	62	1	96.5 ± 2.1	100 ^{+0.0} _{-1.5}	-
	glm-4.7	62	1	96.5 ± 2.1	100 ^{+0.0} _{-1.5}	-
LLM-generated (72 Packages)	Qwen3-max	60	3	93.8 ± 3.0	100 ^{+0.0} _{-1.5}	-
	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



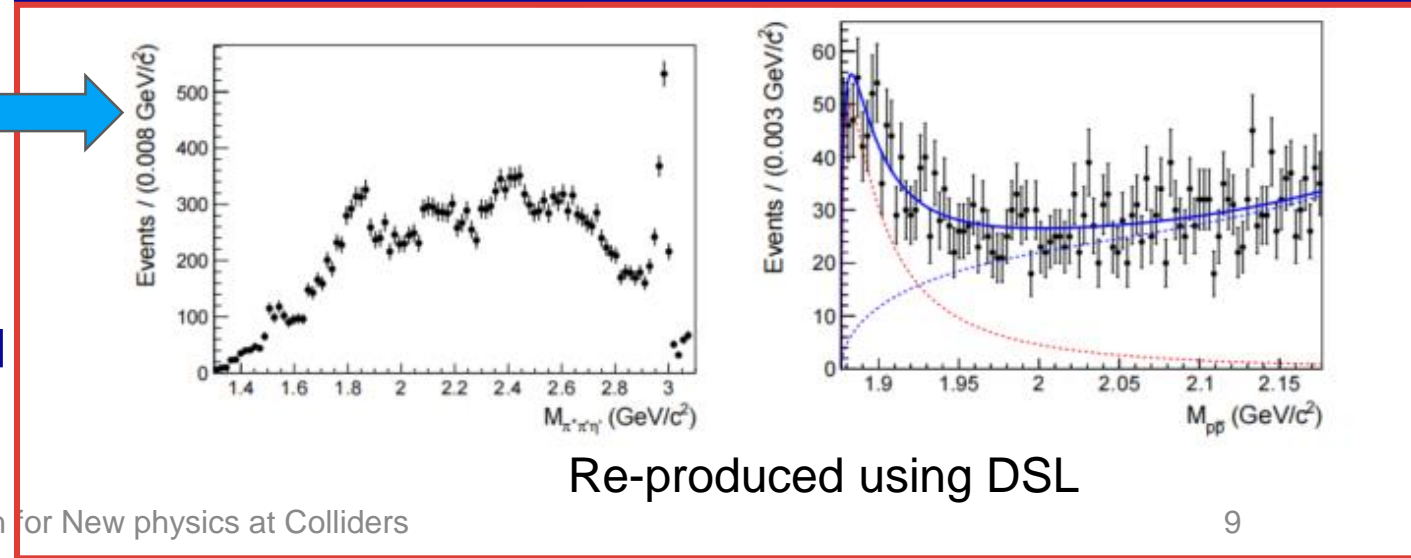
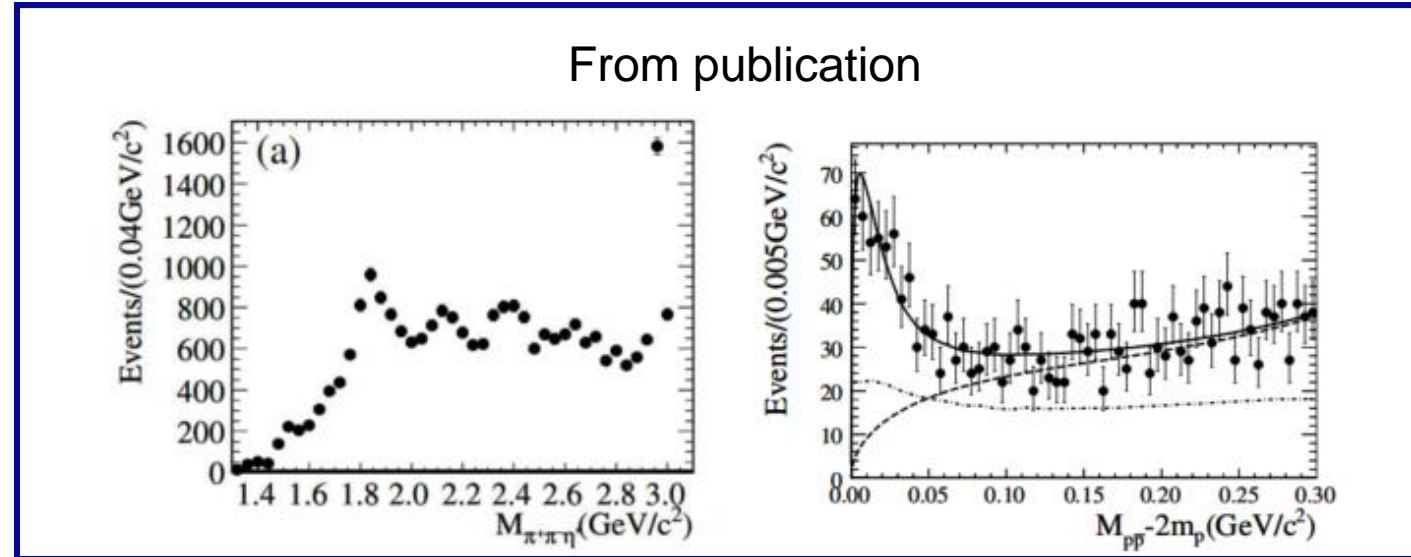
DSL (HepScript) for BESIII analysis

- Necessary for real-world analysis
- A showcase: repeat the previous analysis



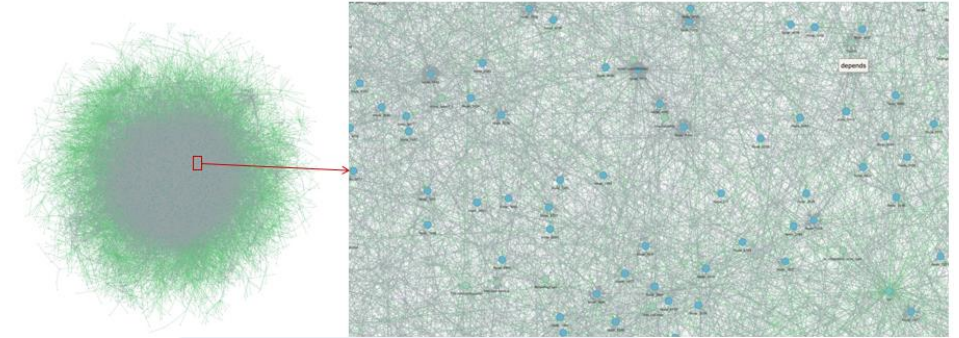
analysis pipeline: simulation, reconstruction, selection, statistical analysis

- Now integrating to Dr.Sai
 - Segmented into skills descriptions and callable toolsets



Code generation using graphRAG

- Total nodes: 47851
- Total edges: 163795
- REPO nodes: 1 | - PACKAGE nodes: 2609 | - FILE nodes: 19347
- TEXTFILE nodes: 5255 | - CLASS nodes: 20639



- Another main challenge in application of agents/LLM
 - Domain code generation
 - Some softwares are not open-source and complex
- We start with open-source software - ROOT
 - Take each class/package as a node
 - Define 3 types of edge
 - contains, extends, depends
 - Construct a large graphRAG
 - Obvious improvement on accuracy
- Now converting to **skills + graphRAG**

Significant improvement

Codegraph_RAG

27.3% to 35%

After five rounds of error feedback

58.33% to 69.33%

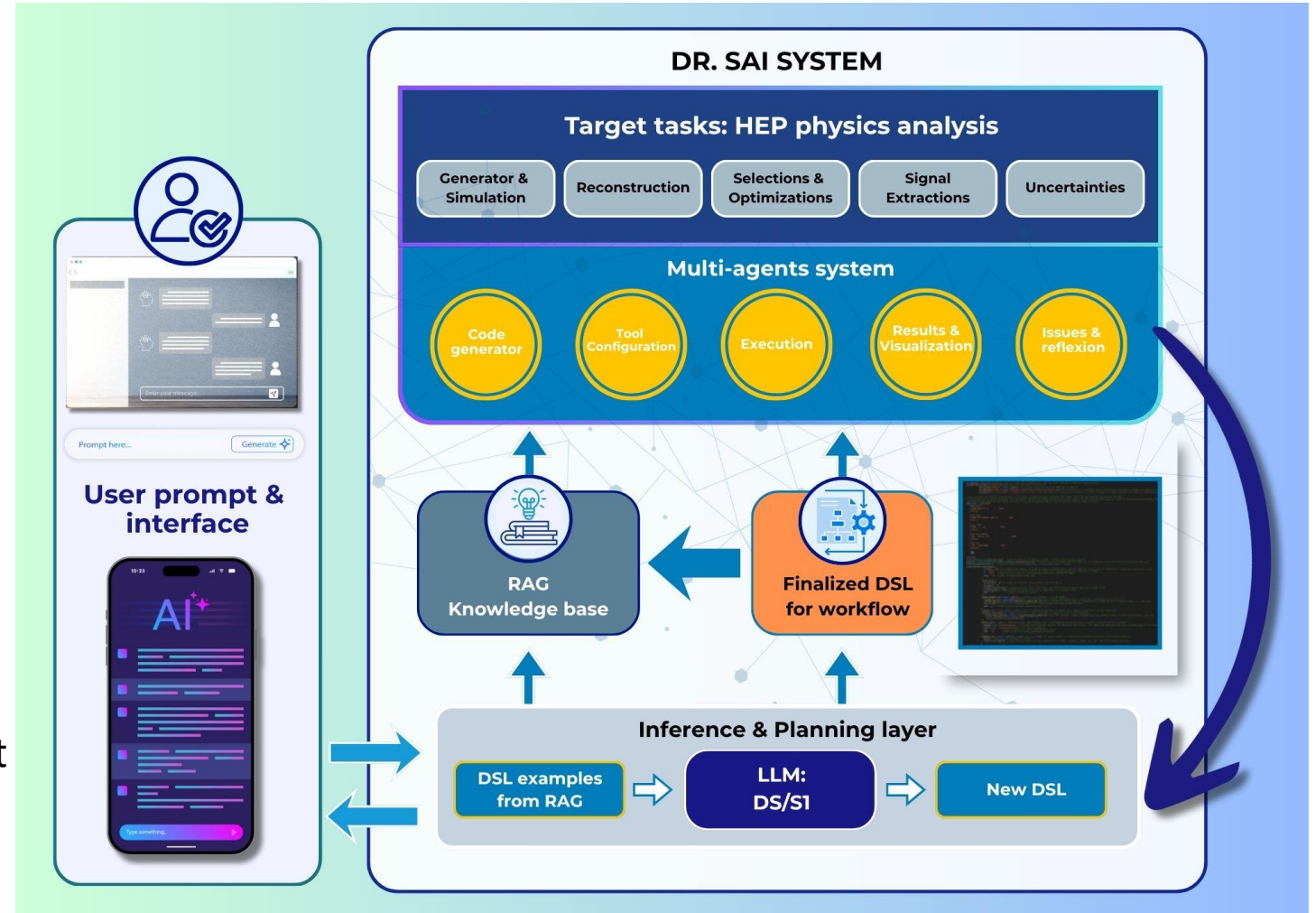
Error_RAG

69.33% to 73.67%

Design tutorials as skills (To be tested)

Next Dr.Sai

- Physicists interact with Dr.Sai using natural language
- The system autonomously invokes tools/code, and executes simple workflows
 - Interactive execution of statistical analysis
- Scientific logic and domain expertise are represented through **DSL/Skills**
 - **Scalable:** Automated pipelines extract from published papers
 - **Continuous Evolution:** revised DSL based on expert feedback
- **Target at 2026 Summer**



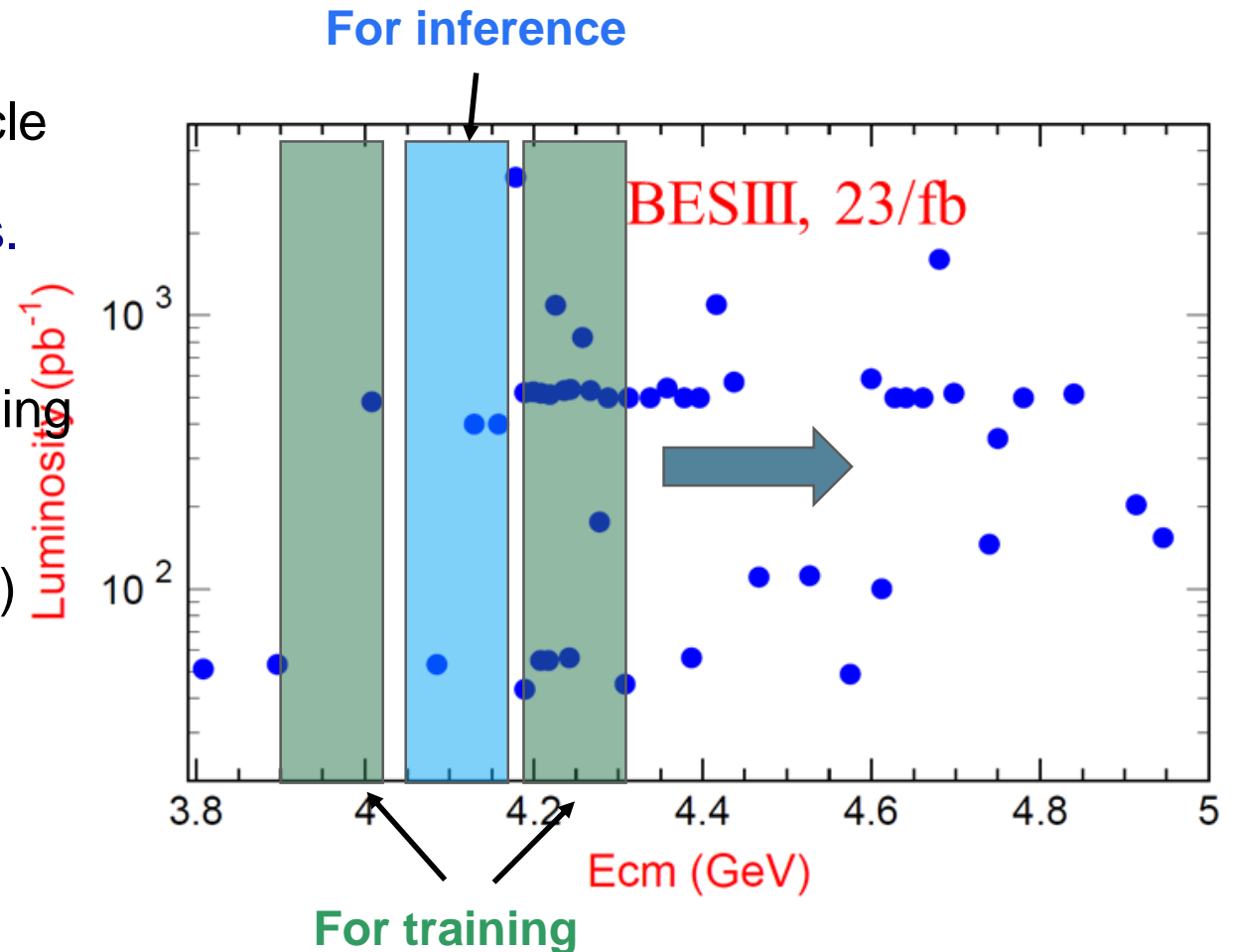
[Front. Phys., 2026, 21\(9\): 096401](#)

What can we do with Dr.Sai

- Systematic search over all possible processes
- On-going study:
 - Test different framework and LLMs, e.g. Claude code, Trae.
 - Measure the cross section exclusively for light hadron processes under 2 GeV
 - Important input for muon g-2
 - Human expert guide Dr.Sai to complete one measurement, then Dr.Sai transfer the Skills and workflow for other processes
 - Please stay tune, the first real results using Dr.Sai is expected at the end of 2026 !

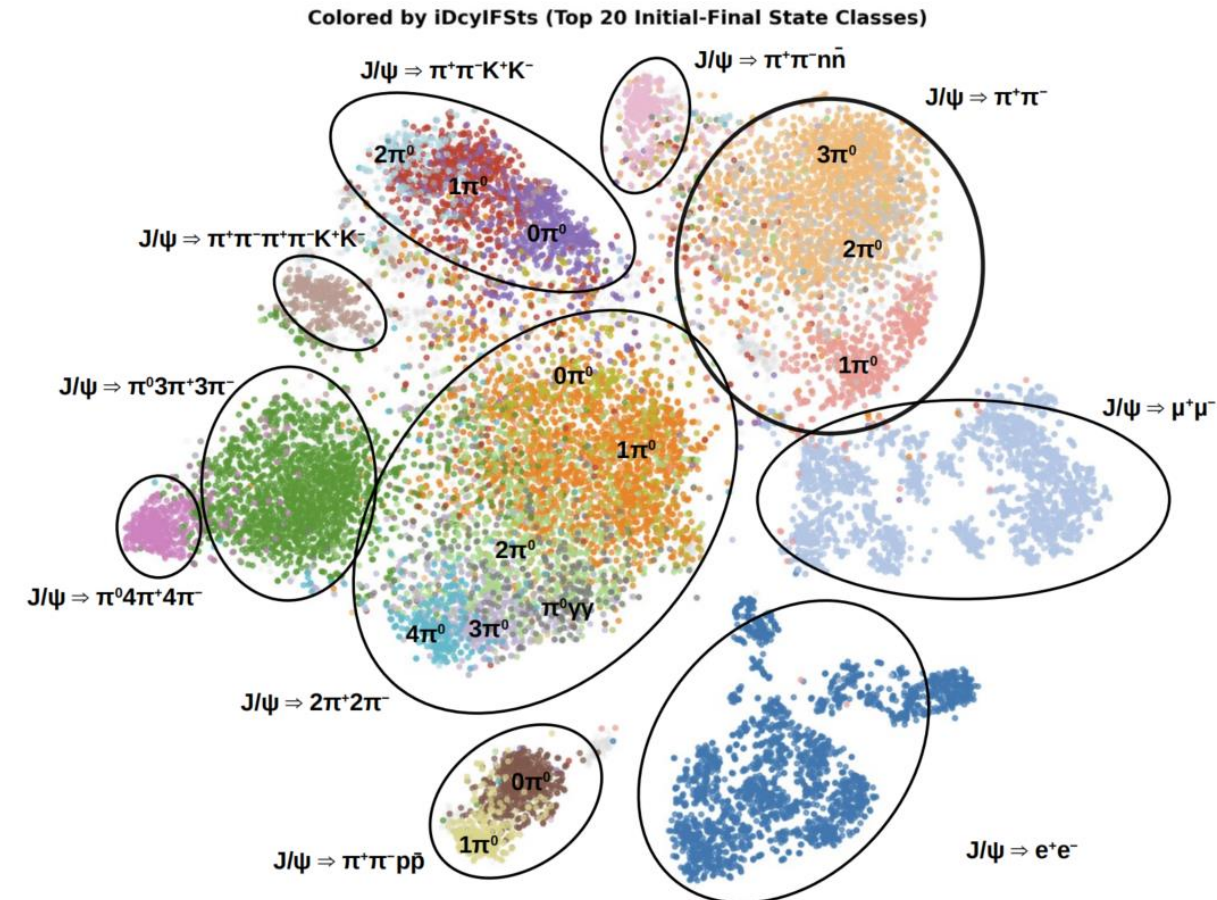
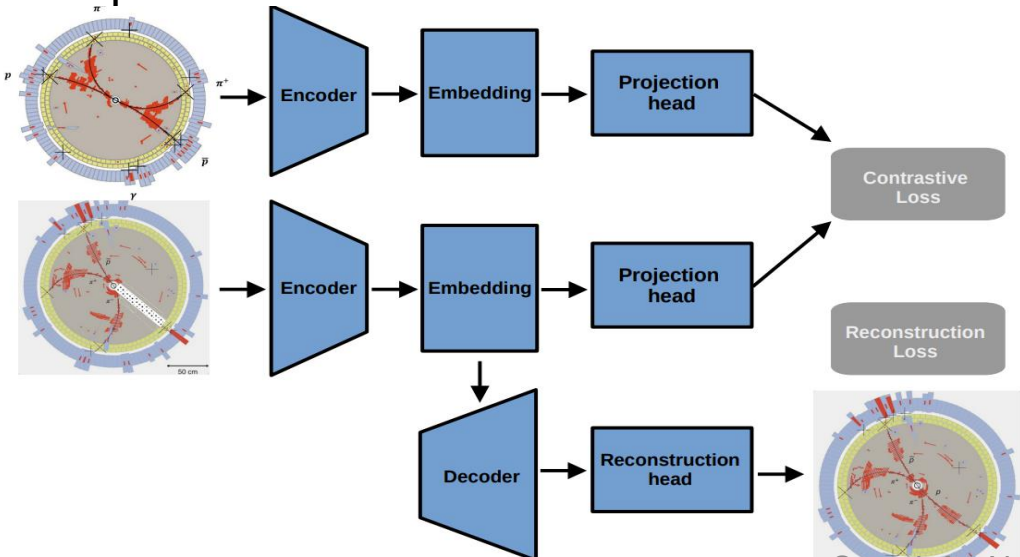
Anomaly detection

- At BESIII, we can do another type of anomaly detection to search for new particle
- We need an AI model robust against c.m.s. energy !
 - Data representation and transfer learning with Foundation Model (FM)
 - Cross-dataset Anomaly Detection (AD) for new hadron/particle



Foundation model for analyses

- 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



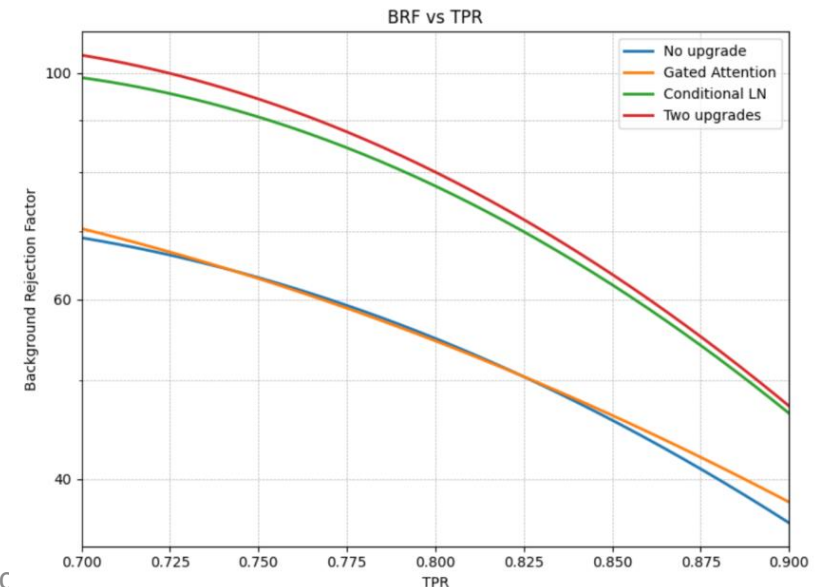
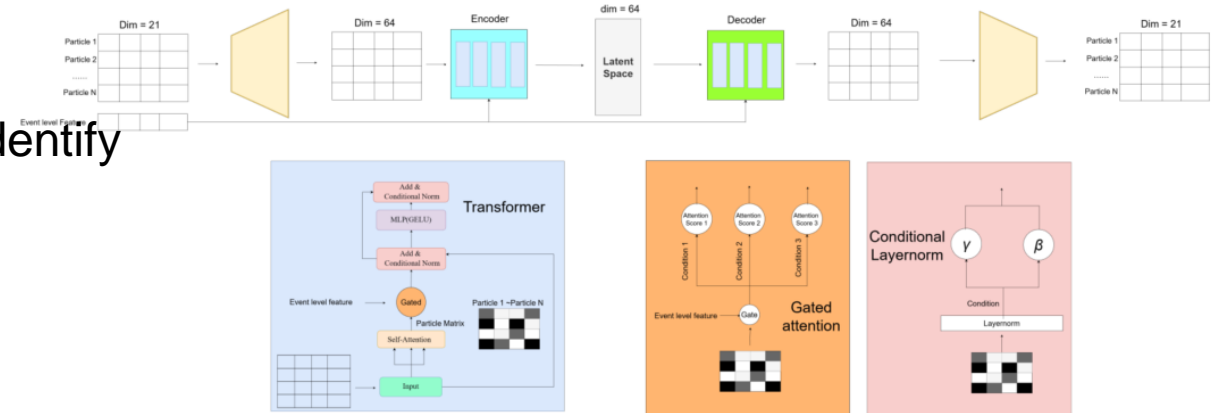
Foundation model for analyses

- A showcase: classification for 99 classes
- Better recall when precision are the same w.r.t. traditional method with kinematic constrains
- Next: add momentum conservation into the model for further improvement

Decay Channel	Foundation Model			Traditional Method		
	Precision	Recall	F1-score	Precision	Recall	F1-score
$J/\psi \rightarrow \pi^+ \rho^-, \rho^- \rightarrow \pi^0 \pi^-$	0.85	0.88	0.87	0.72	0.74	0.73
$J/\psi \rightarrow \pi^- \rho^+, \rho^+ \rightarrow \pi^0 \pi^+$	0.83	0.90	0.86	0.71	0.74	0.72
$J/\psi \rightarrow \pi^0 \rho^0, \rho^0 \rightarrow \pi^+ \pi^-$	0.87	0.92	0.90	0.82	0.82	0.82
$J/\psi \rightarrow \pi^+ \pi^- \pi^0$	0.74	0.76	0.74	0.74	0.70	0.71

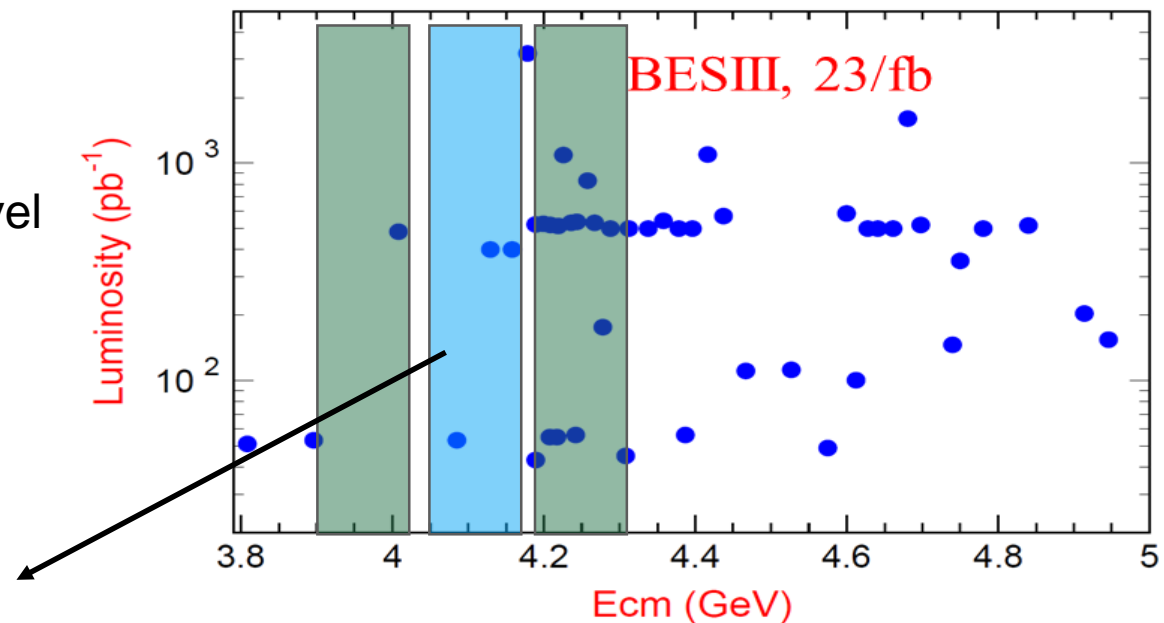
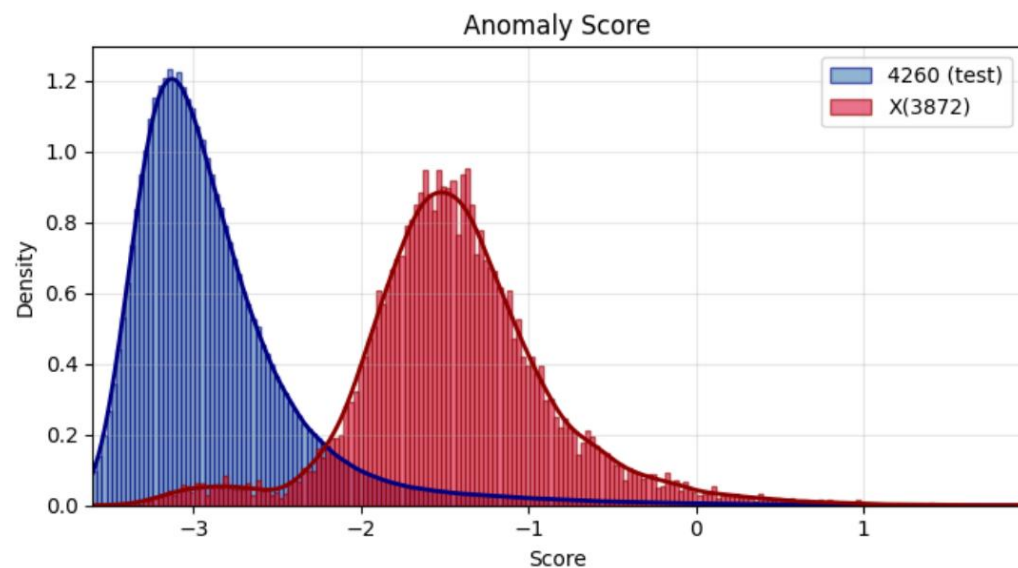
Anomaly detection

- Motivation:
 - Many datasets at different c.m.s. energy
 - We can use one dataset to train a model then identify the anomaly events in another dataset
- Model: Self-attention + AutoEncoder
 - Event-level features(energy, etc.) as conditions
- Methodology and Evaluation:
 - training sample: fake data + inclusive MC at 4.23 GeV + inclusive MC at 4.26 GeV
 - Inference on fake data + signal MC at 4.26 GeV
 - Preliminary study shows good robustness against c.m.s. energy !
 - Obvious improvement by introducing gated attention and conditional layernorm



What can we do with FM+AD

- Global search for new particle at BESIII
 - Similar to direct comparison at dataset level
 - Sensitive to new hadrons with different mass/widths
- Next: Combine FM and AD



Why we need it?

- It trained on real data without further selections
 - Expected to be sensitive to most of the new particles
 - We can release the ML model weights to public, theorists can test NP model easily

Summary and outlook

- Possible two new approaches to search for new physics at lepton collider
 - Unique scenario w.r.t. hadron collider
 - Systematic scan of all possible process
 - Autonomous analysis using Dr.Sai - working on real data now
 - Global search using cross dataset anomaly detection - finalizing prototype
 - A comprehensive representation of event from FM
 - Cross-dataset anomaly detection with robustness against c.m.s. energy
- Several studies are in the internal review, soon to be released
- Collaborations are more than welcome !

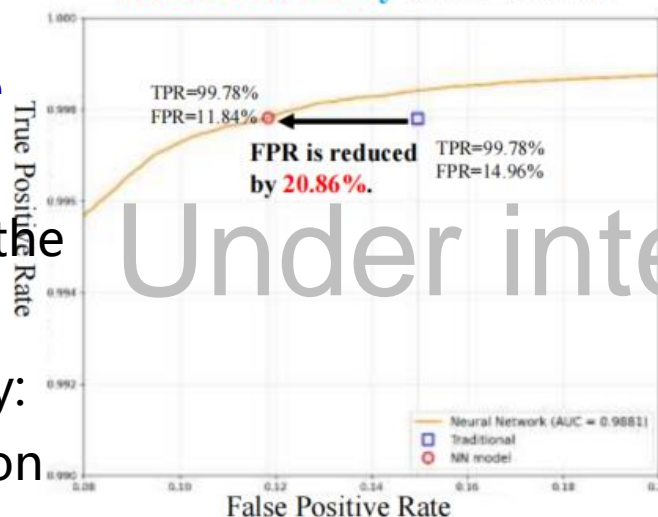
backup

Trigger: test new strategy at lepton collider

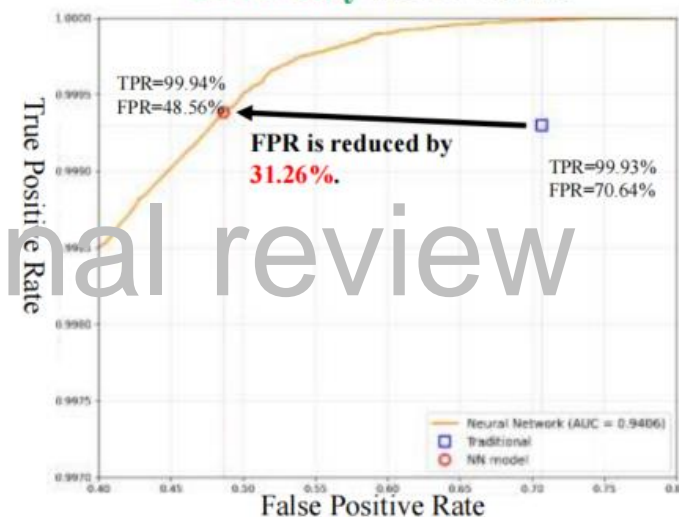
- At lepton collider, trigger rate is dominant by beam induced background
- Traditional method can achieve >99% efficiency for hadron events
- Background suppression via light NN
 - Effectively reduce background rate
 - Preserves hadron efficiency
- After pruning and migrating to FPGA, the latency is <300ns
- Next we can test a new trigger strategy:
 - instead of selecting events based on signal signature, we can select and remove beam background

	Num of Event	Num of Hadron	Hadron Ratio
J/ψ peak	8.74M	2.58M	29.47%
$\sqrt{s} = 4.680\text{GeV}$	12.75M	0.27M	2.1%

MDC&TOF only ROC Curve



EMC only ROC Curve



Under internal review