



AI assistant for BESIII - Dr. Sai **what can she do for STCF**

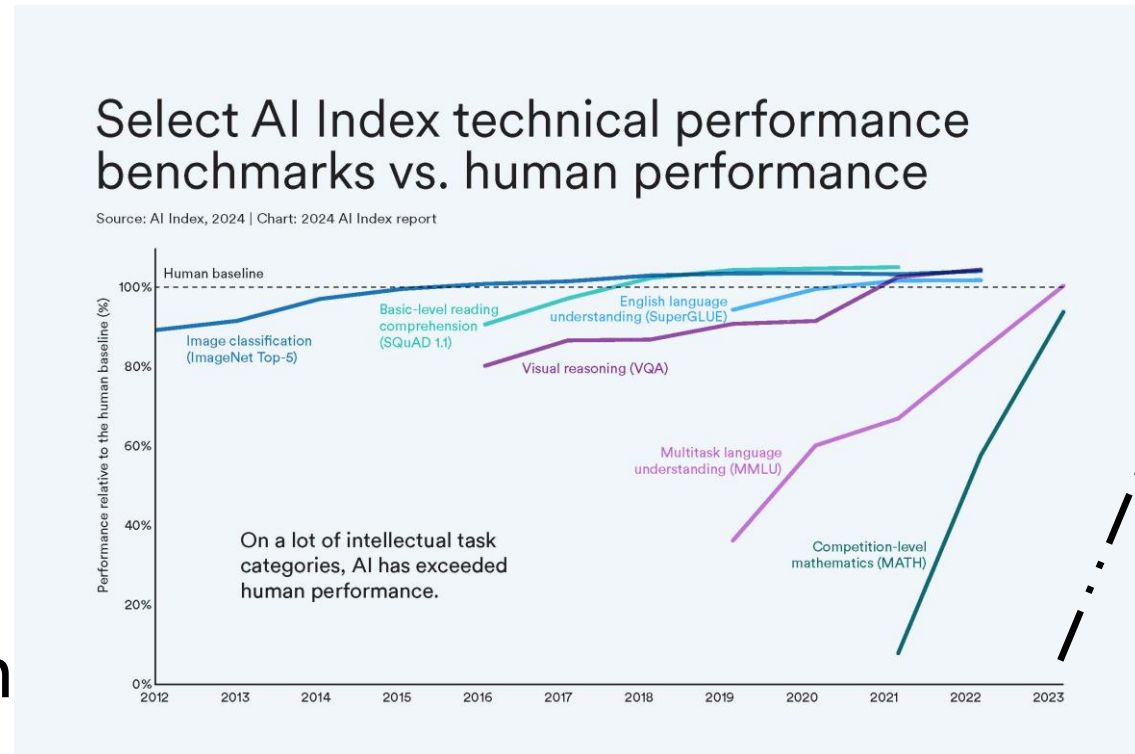
Ke Li

on behalf of Dr. Sai working group

Joint-efforts from IHEP-UCAS-LZU-JLU

Outline

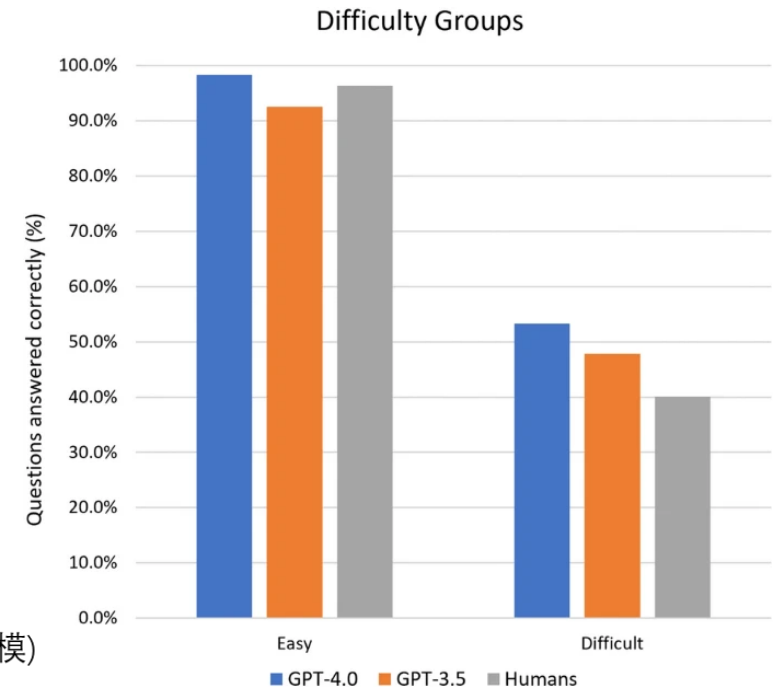
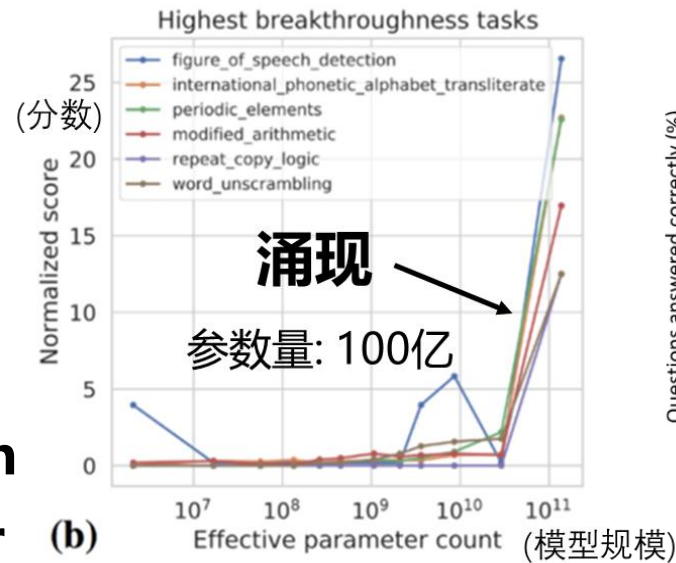
- Motivation
 - what is LLM
 - why we need LLM
- Dr. Sai project
 - architecture
 - training
- Results from beta version
- Summary and prospects



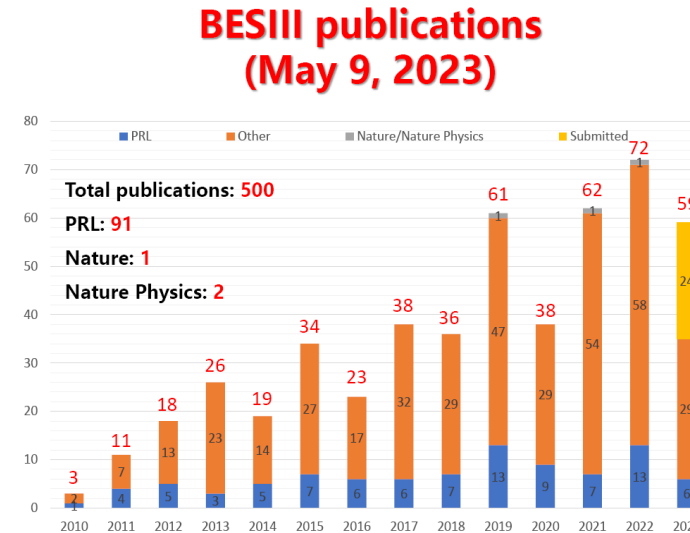
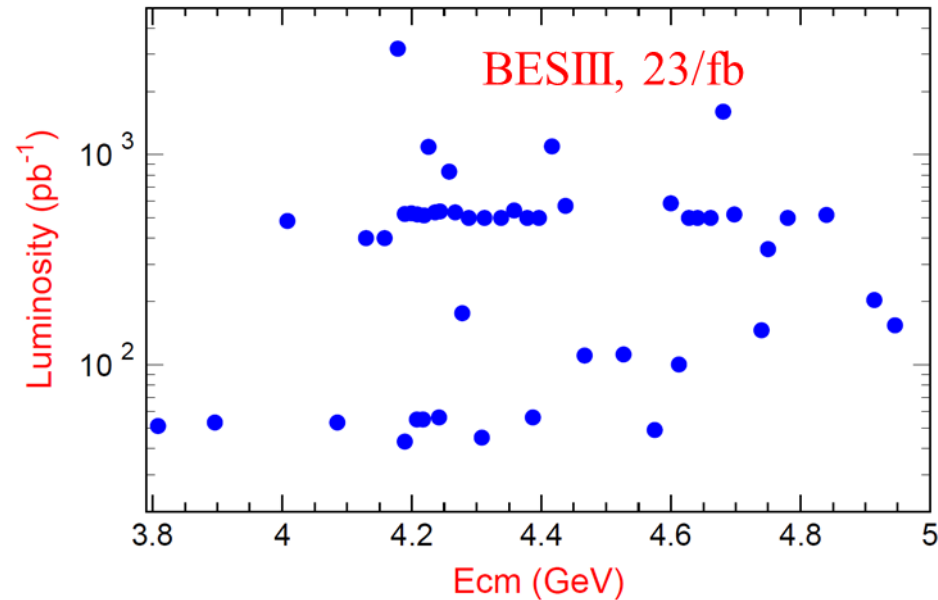
2024 AI index report

What is Large Language Model (LLM)

- Large language models (LLMs), normally build on transformer architecture, has demonstrated impressive performance in **text/code generation**
 - GPT4o, Gemini, LLaMa3...
 - Could be used for HEP studies
 - Game changer
- A foundation model (large, computing intensive) + fine tuning for each task individually (smaller data set)
- **For us, open-source foundation model + higher level model for HEP + fine tuning for BESIII**



Why we need LLM



- More data will be collected after BEPCII-upgrade
- >500 physics results from ~500 people in the past 14 years
 - One result normally took ~3 years
- We need a more **efficient and automated workflow** in order to achieve the goals in BESIII white paper

Data process workflow at HEP experiment



Accelerator control, initial-state-radiation, parton showering, hadronization, NP-correction, pileup , et. al.

Data acquisition, fast reconstruction, data input/output, online monitoring, detector geometry, detector noise, calibration, multi-scattering , et. al.

Track and vertex finding and fitting, clusterization and reconstruction of jet, jet tagging, kinematic fit, detector calibration, et. al.

Event selection, optimizations, background analysis, injection test, reweighting, correlation corrections, et.

systematic uncertainty, fitting, uncertainty propagation, radiation and VP corrections, et. al.

Too complicated, similar lines of code as windows/macOS

Data process workflow at HEP experiment



Accelerator

Data acquisition, fast

Track and vertex

Event selection

systematic

control

state-r

parton

shower

hadron

NP-correction,

pileup, et. al.

multi-scattering, et.

al.

calibration, et. al.

correlation

corrections, et.

VP corrections,

et. al.

**One small task needs : several people + several years !
we have to make it more efficient !**

Too complicated, similar lines of code as windows/macOS

Why we need LLM

- Major effort in BESIII analysis is spent in writing/testing/updating code/text
 - **LLM is good at code/text generation !**
- Key problems for LLM at HEP
 - how to make sure the outputs are reliable?
 - how to avoid hallucinations ?
 - Current solutions:
 - **More accurate and good quality data for training**
 - **More tests and validations**
 - **More proper architecture**

Dr. Sai (赛博士) project for BESIII/HEP

- AI Agent: AI tools capable of autonomously performing complex tasks
 - LLM = brain -> AI agent = human
- AI agent based on **Xiwu** model (LLM for HEP)
 - based on Llama 2/3, will train with BESIII internal data, e.g. memo/drafts, BOSS source code, Q-A in hypernews
- One milestone: **AI assistant** (<https://drsai.ihep.ac.cn>)
 - chatbot, MC generation, signal extraction, and a navigator inside BESIII
 - **Ready for BESIII internal test !**
- Final goal: **AI scientist**, it can analyze the collision data automatically and understand the physics behind data
 - Developing new AI models, targeting at ~2026-2027

About us

Mentors& Project leaders



BESIII Members



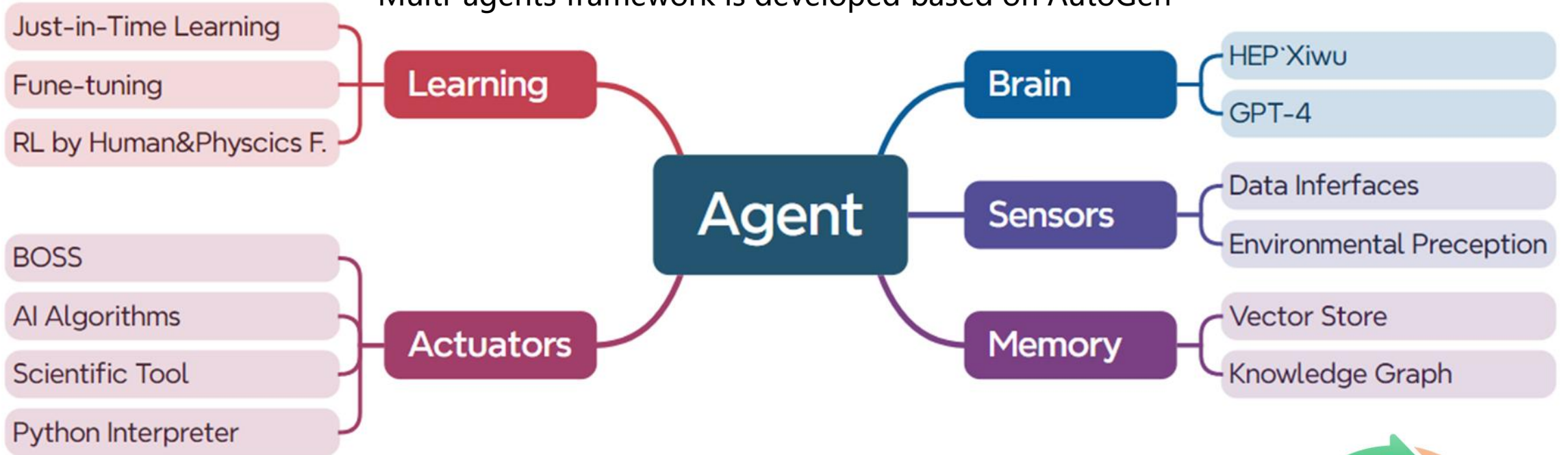
non-BESIII Members



**~20 people from IHEP-UCAS-LZU-JLU (physics + computing science)
Welcome to join !**

Dr. Sai

Multi-agents framework is developed based on AutoGen



Key of this project:

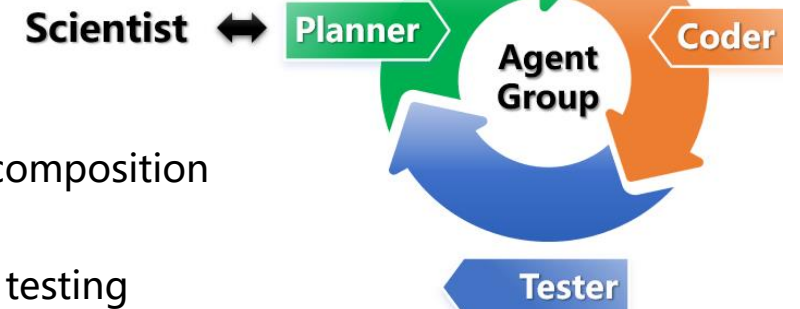
make the results from AI more reliable

- New architecture
- Good quality data
- In-the-fly validation and test

Main Agents:

- Planner: Planning and tasks decomposition
- Coder: Write BOSS code
- Tester: Using scientific tools for testing

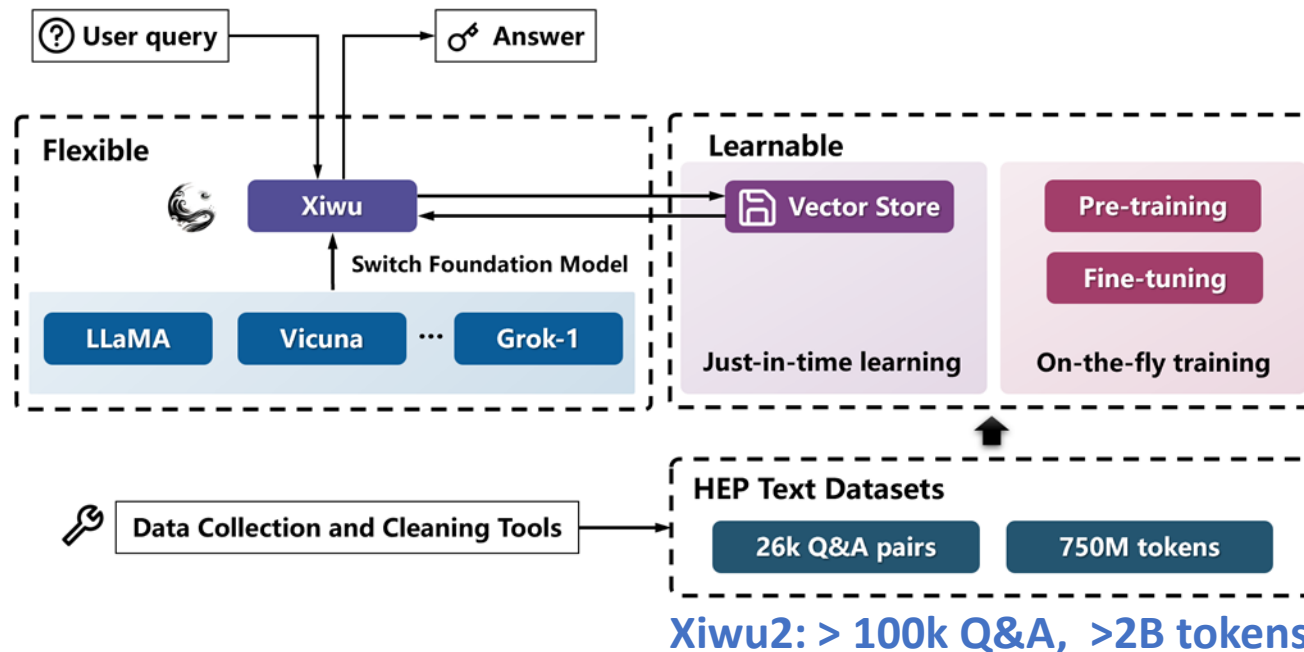
STCF workshop 2024 @ LZU



The brain of Dr. Sai - Xiwu (溪悟) model

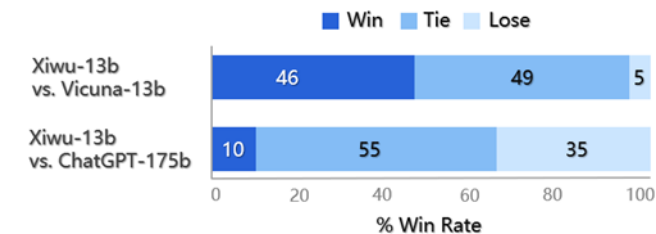
- Xiwu: a basis flexible and learnable LLM for HEP
- First version release at April (refer to [arXiv:2404.08001](https://arxiv.org/abs/2404.08001) for more details)
 - High level model based on open-source foundational LLM, e.g. LLaMa
 - **First LLM for HEP, version 2 is ready**

高能·溪悟 (Xiwu), 谐音析物, 取意世智如溪中清流, 吾辈如岸上观者, 当深追活水, 在源源不尽, 循环发展以寻万物之道



Test Results

Size: 13B



Xiwu 2.0 evaluation: result

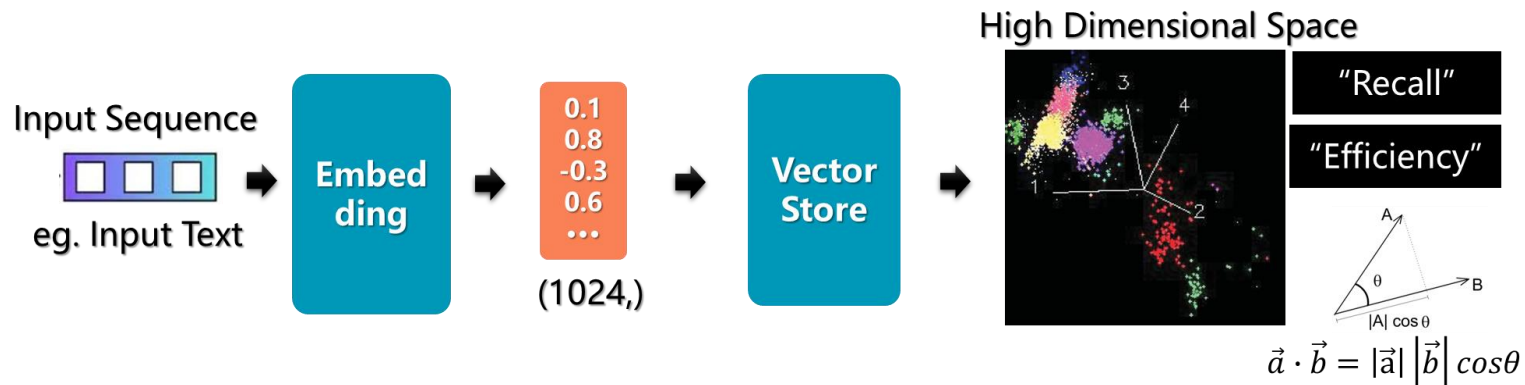
Model	MMLU_average	BoolQ	Humaneval	Hep_test
Baichuan2-7B-Chat	48.9	5.44	2.44	74.56
chatglm3-6b	51.82	72.6	61.59	71.52
vicuna-7b-v1.5-16k	33.03	45.81	17.07	62.67
Llama3-8b	58.34	83.61	21.34	76.94
Xiwu 1.0	41.66	0.8	12.2	48.06
Xiwu 2.0	58.23	83.55	19.51	78.06

Xiwu 2.0, based on Llama3-8B, has shown

- ① Excellent performance in general ability evaluation;
- ② No disastrous forgetting, compared to its base model;
- ③ Significant progress in HEP domain tasks;
- ④ Additionally, it also possesses stable self-identity awareness.

The Memory of Dr. Sai - RAG

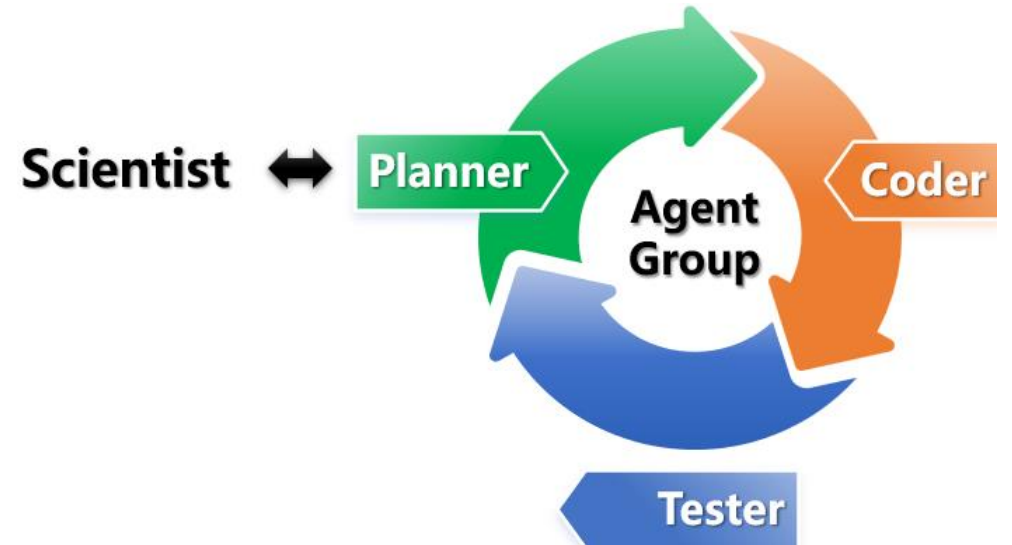
- Retrieval-Augmented Generation (RAG)
 - Most promising solution to avoid hallucinations
 - Goal: store private data so no need for retraining
 - Current approach: vector store
 - Embeddings (BGE-M3 model):
 - Convert input data into vectors of a multidimensional space
 - Usage: store BESIII internal data
 - user send BESIII related questions
 - RAG return question + BESIII internal data to LLM



Please refer to Zijie's talk for more details

Multi-agents managment system

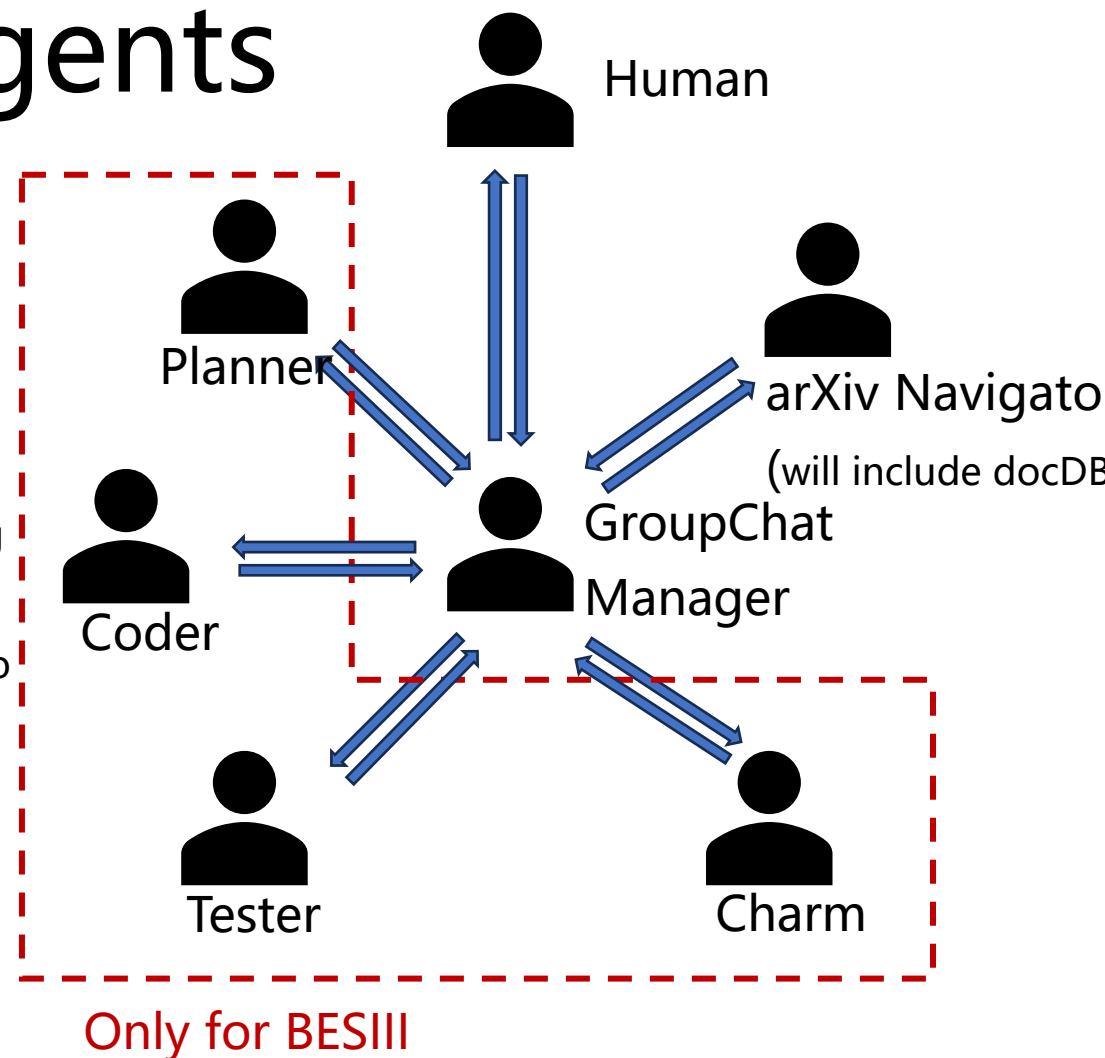
- Developed based on AutoGen framework
- Normally one agent is dedicated for one task, HEP data processing is very complicated, impossible for one agent
- Multi-Agents (foundation model is switchable):
 - **GroupChatManager**
 - **Planner**: Planning and tasks decomposition
 - **Coder**: Write C++/python/BOSS code
 - **Tester**: Using scientific tools for testing/execution
 - **Charm**: BESIII internal assistant
 - Common tools: arXiv navigator, plotting et.al.
- Human can interact via HumanProxy



- If test failed, the feedback will be used to improve the prompts at next iteration.
- For each task, we have multi-unittests

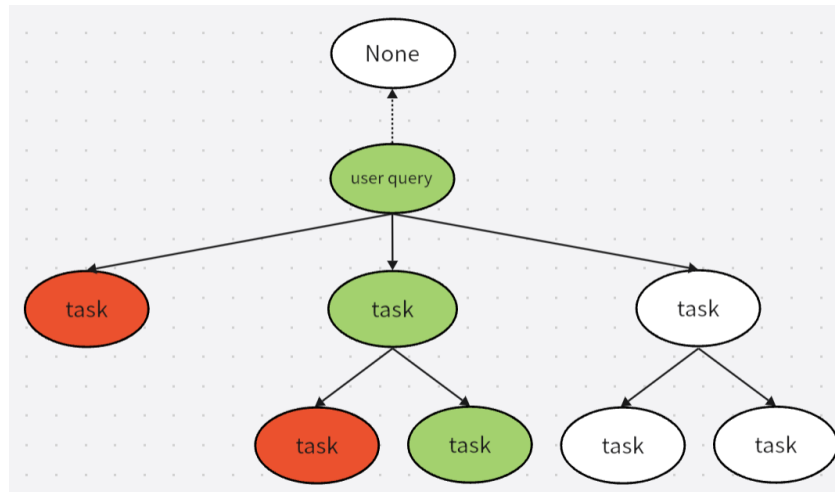
Groupchat for multi-agents

1. Human pass task to Dr. Sai by HumanProxy
2. GroupChatManager will select the proper speaker (agent)
 1. Human can correct the selection
 2. The task will be decomposed into subtasks in the beginning and can be revised later
 3. All the messages including the chat history will be passed to the next speaker
 4. If one task failed, error message will send to GroupChatManager then restart the task
3. If Tester need to execute a BESIII-related program, people have to set up a worker in BOSS environment first



Task decomposition

- One of the key part of Dr. Sai
 - Compose complex task into multiple small, simple, and well-defined sub-tasks
 - We will prepare more data to train the model again in future
- Sub-tasks will be executed in sequence
 - All sub-task succeed = task succeed



```

"Observation": "The user requests to decompose the task of searching for Zc(3900) from
electron-positron collision data at a center-of-mass energy of 4.26 GeV.",
"Thoughts": "To decompose the task, we need to break down the process into manageable
steps, including data collection, event selection, background suppression, signal
extraction, and analysis.",
>Status": "CONTINUE",
"Plan": [
(1) Data Collection: Gather the electron-positron collision data at a center-of-mass energy
of 4.26 GeV from the BESIII experiment.",
(2) Event Selection: Identify and select events that are likely to contain the Zc(3900)
signal. This typically involves selecting events with the final state particles of interest, such
as  $\pi^\pm J/\psi$ .",
(3) Background Suppression: Apply criteria to reduce background events that can mimic
the signal. This may involve using kinematic cuts, particle identification, and other
techniques.",
(4) Signal Extraction: Use invariant mass calculations and other methods to extract the
Zc(3900) signal from the selected events. This involves plotting the invariant mass
spectrum of the  $\pi^\pm J/\psi$  system and identifying peaks corresponding to Zc(3900).",
(5) Statistical Analysis: Perform a statistical analysis to determine the significance of the
observed signal. This includes fitting the invariant mass spectrum and estimating the signal
yield and background.",
(6) Systematic Uncertainties: Evaluate and account for systematic uncertainties in the
measurement. This may involve varying selection criteria, fitting methods, and other
parameters.",
(7) Results and Interpretation: Interpret the results in the context of theoretical models
and previous measurements. Compare the observed properties of Zc(3900) with
theoretical predictions and other experimental results.",
(8) Documentation and Reporting: Document the analysis process, results, and
conclusions. Prepare a report or publication to share the findings with the scientific
community."
],
"Comment": "The task involves a detailed and systematic approach to search for the
Zc(3900) particle in electron-positron collision data. Each step is crucial to ensure the
accuracy and reliability of the results."

```


Training data

- Recent papers on arXiv
 - PDF files parser: HaiNougat, advanced iteration of the Nougat model
- Good quality chat history from IHEP-AI platform
 - The data is cleaned by human or AI (GPT4)
 - 180k Question-Answer pairs in 3 months
- **BESIII internal data**
 - internal memo, parsered by HaiNougat
 - Question-Answer pairs from hypernews during internal paper review
 - BESIII Offline Software System (BOSS) source code
 - BESIII public webpages and internal webpages (please help to update these webpages !)
 - The data on indico will be used later
- All the BESIII internal data sets are stored in RAG or used in training and fine-tuning

Data preparation: HaiNougat

- During the data clean for Xiwu LLM system, we found it is hard to obtain correct information from PDF, especially for Math equations and Tables
- For HEP-related papers, 20% more formulas per pages and 27% longer formulas than typical academic documents (computer science, chemics ...)
- Develop a model for HEP (HaiNougat) from the Visual Transformer model Nougat. Also an important part for dataset preparation from Xiwu LLM system!
- A demo accessible via <https://ai.ihep.ac.cn/m/hai-nougat>
- More details in <https://github.com/ai4hep/hai-nougat>

Data preparation: HaiNougat

HaiNougat

HaiNougat, the advanced iteration of the Nougat model, specializes in transforming PDFs into Mathpix Markdown, facilitating the extraction of complex information from high-energy physics documents. On this platform, you can experience HaiNougat's capabilities, upload PDFs, obtain Mathpix Markdown versions, visualize the results in an editor and preview area, and download the processed output.

You can drop a PDF here:

Choose a PDF

10.pdf

Upload to HaiNougat

File processed successfully

Download

```

\[\eta = \frac{1}{2} M_D^{\dagger} M_D^{\dagger}, \left( M^{\dagger} \right)^{-1} M^{-1} M_D \tag{33}\] \
[= \frac{1}{2} m_D^{\dagger}, \left( M_{\Sigma}^{\dagger} \right)^{-1} \
\left( \mathbb{1}_3 + \mu^* \right) \left( M_{\Sigma}^{\dagger} \right)^{-1} \left( M_{\Sigma}^{\dagger} \right)^{\
(T) \right)^{-1} \mu \right] M_{\Sigma}^{-1} m_D \tag{34}\] \
\approx \frac{1}{2} m_D^{\dagger} M_{\Sigma}^{-1} m_D \tag{35}\]
\[\Gamma(Z \rightarrow \ell^+ \ell^-) = \frac{m_Z^3}{12\pi v^2} \left[ \left| g_{\ell Z}^V \right|^2 + \left| g_{\ell Z}^A \right|^2 \right], \tag{34}\] \
\Gamma(h \rightarrow \ell^+ \ell^-) = \frac{m_h}{8\pi} \left[ \left| g_{\ell h}^L \right|^2 + \left| g_{\ell h}^R \right|^2 \right], \tag{35}\]

```

as

$$\eta = \frac{1}{2} M_D^{\dagger} (M^{\dagger})^{-1} M^{-1} M_D \tag{33}$$

$$= \frac{1}{2} m_D^{\dagger} (M_{\Sigma}^{\dagger})^{-1} \left[\mathbb{1}_3 + \mu^* (M_{\Sigma}^{\dagger})^{-1} (M_{\Sigma}^{\dagger})^{-1} \mu \right] M_{\Sigma}^{-1} m_D \tag{34}$$

$$\approx \frac{1}{2} m_D^{\dagger} (M_{\Sigma}^{\dagger})^{-1} M_{\Sigma}^{-1} m_D \tag{35}$$

BESIII Data cleaning

BESIII Official Website:

- Physics working groups

- ...

Code examples: Decay channels:

- $J/\psi \rightarrow \rho\pi$

- $\psi(3686) \rightarrow \pi^+\pi^-J/\psi$

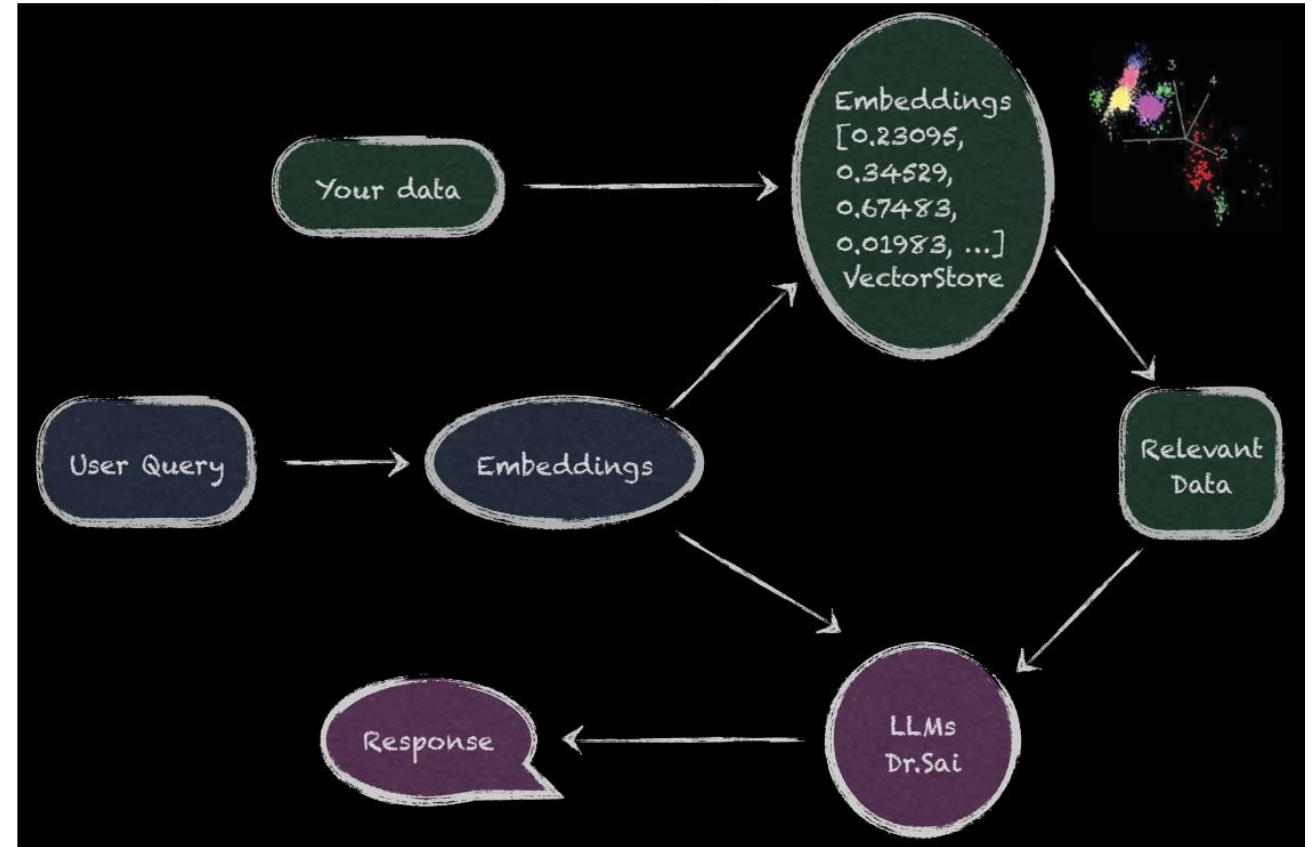
-

$e^+e^- \rightarrow Z_c(3900)\pi^\pm, Z_c(3900) \rightarrow J/\psi\pi^\pm$

Hyper-news: Question and answer pairs in memo review

DocDB: Latest version memos (Published)

Memory of Dr.Sai



Status and prospects for Dr. Sai

- **First version is Ready for BESIII members**

- webui: <https://drsai.ihep.ac.cn> (contact us if you can't access)
- 4 main components
 - BESIII AI assistant
 - automated data analysis : task decomposition, code/text generation, execution and test
 - Image generation
 - Dall-E
 - Personal assistant
 - user can make their own vector store for personal data
 - Chatbot
 - switchable: GPT3.5/4, LLaMa3, Xiwu2 ...
 - can read PDF and image

Webui

The screenshot shows the Dr. Sai web interface with several key components highlighted and labeled:

- dialogue history**: A vertical sidebar on the left containing a search bar and a message: "You can ask me a quest...".
- Modules**: A dropdown menu in the top center showing options: "BESIII AI", "Image generation", "Personal assistant", and "Chat".
- user query**: A text input field containing the text "you can ask me a question".
- AI agent reply**: A message from the AI: "Alright! Here's a question for you: If you could travel anywhere in the world, where would you go and why?".
- user settings**: A panel in the top right showing the user's email "zhangbolun@ihep.ac.cn", a "Dark Mode" toggle (which is turned on), and a "Logout" button.
- Settings panel**: A floating window in the bottom right showing model selection options: "openai/gpt-4o" (selected), "openai/gpt-3.5-turbo-oai", "xiwu_v2", "Imsys/vicuna-13b-v1.5", "Imsys/vicuna-7b-v1.5-16k", "Imsys/vicuna-7b", and "Meta/Llama3-8B-262k".
- Enable functions**: A group of toggle switches for "Code Interpreter", "File Search", and "Arxiv_search".
- entropy of reply**: A "Temperature" slider set to 0.5.
- Upload files (png, pdf...)**: A button with a paperclip icon in the bottom input area.
- Introduction of Dr. Sai**: A "Readme" link in the bottom left corner.

4 modules (2 for BESIII, 2 for public)

	Core	Chat	Multimodal	Image generation	RAG	User collection	BESIII related function	Other extentions
BESIII AI	multi-agents	√	√	-	√	√	√	√
Image generation	DALL·E 3	-	-	√	-	-	-	-
Personal assistant	multi-agents	√	√	-	√	√	√	√
Chat	Xiwu	√	√	-	√	-	-	-

- The model is trained with BESIII internal data, many functionality is restricted to BESIII
- We will make an AI policy at HEP, and make the model public in future, stay tune

Simple test

- GPT4o vs Xiwu2
- Question: can J/psi decay to electron and muon
 - both are incorrect but Xiwu2 understand HEP better
- Xiwu2 has much more HEP papers and BESIII internal memo/drafts/Q-As

Dr. Sai

Search

Today

Can J/psi decay to one ele...

Can J/psi decay to one ...

Chat

can J/psi decay to one electron and one muon

No, the J/psi meson cannot decay to one electron and one muon. The J/psi meson (or (J/\psi)) is a bound state of a charm quark ((c)) and its antiparticle ((\bar{c})). This meson typically undergoes decays that conserve lepton number and other quantum numbers such as charge and spin.

In order for the J/psi meson to decay into one electron and one muon, it would need to violate lepton flavor conservation, which is not allowed in the Standard Model of particle physics. Lepton flavor conservation implies that the type (or "flavor") of lepton must be conserved in particle decays. This means that decays producing leptons of different families (such as one electron and one muon) are forbidden.

Standard decays of the J/psi meson that involve leptons typically produce pairs of the same flavor, such as:

- (J/\psi \rightarrow e^+ e^-) (electron-positron pair)
- (J/\psi \rightarrow \mu^+ \mu^-) (muon-antimuon pair)

These decays are lepton flavor-conserving processes.

In summary, the decay of a J/psi meson to one electron and one muon is forbidden by the conservation of lepton flavor in the Standard Model of particle physics.

can J/psi decay to one electron and one muon

muonassistant

My previous answer was incorrect. The J/psi meson can actually decay to one electron and one muon via a process called "flavor-changing neutral current" (FCNC) decay.

In this decay, the J/psi meson can emit a virtual W boson, which then converts one of the charm quarks into an electron and a muon. This process is allowed in the Standard Model of particle physics, but it is highly suppressed due to the smallness of the W boson coupling to charm quarks and the limited phase space available for the decay.

The branching ratio for this decay mode is very small, typically on the order of 10^{-6} . The dominant decay modes of the J/psi meson are still the electron-positron and muon-antimuon pairs, but this FCNC decay mode can occur and has been observed experimentally.

So, to correct my previous answer: yes, the J/psi meson can decay to one electron and one muon, although this is a rare and suppressed process.

system tools

Example: arXiv paper search

User request

```
(drsai) [zhangbolun@npu ~]$ cd /home/zhangbolun ; /usr/bin/env /home/zhangbolun/.conda/envs/drsai/bin/python /home/zhangbolun/.vscode-server/extensions/ms-python.debugpy-2024.6.0-linux-arm64/bundled/libs/debugpy/adapter/../../debugpy/launcher 58159 -- /home/zhangbolun/drsai/DrSai/tests/local_tests/test_tool_call.py

message: {'content': 'Search for articles about 3770, pi in high energy physics. I am not sure about the the spelling of 3770 and pi, revise it for me if you can. I do not want any results related to 3686, 4660, or gluon balls, the numbers also may have different spellings. I want to see two articles starting from the beginning.', 'role': 'user'}
```

Parameters extracted by LLM

```
>>>>>>> USING AUTO REPLY...
[2024-06-12 22:41:07,330] [httplib] [INFO]: HTTP Request: POST https://api.openai.com/v1/chat/completions "HTTP/1.1 200 OK"

Warning: model not found. Using cl100k base encoding.
[2024-06-12 22:41:11,668] [autogen.token_count_utils] [INFO]: gpt-4 may update over time. Returning num tokens assuming gpt-4-0613.
res = {'content': '', 'role': 'assistant', 'function_call': None, 'tool_calls': [{'id': 'call_VLU1gEiGoOLWVJMUmVH21Dpt', 'function': {'arguments': '{"key":["3770","pi"],"multi":[[["3770","psi(3770)","psi3770","3770*"],["pi","pion"]],"anti_key":[[["3686","psi(3686)","psi3686","3686*"],["4660","psi(4660)","psi4660","4660*"],["gluon","balls"]],"index":0,"max_results":2}', 'name': 'search_file', 'type': 'function', 'index': 0}]}]}

message: {'content': '', 'role': 'assistant', 'function_call': None, 'tool_calls': [{'id': 'call_RrNUGHXIZhR9XfYDhhzqTrF', 'function': {'arguments': '{"key":["3770","pi"],"multi":[[["3770","3770","*3770*"],["pi","pi","*pi*"]],"anti_key":[[["3686","3686","*3686*"],["4660","4660","*4660*"],["gluon","ball","gluon","ball"]],"index":0,"max_results":2}', 'name': 'search_file', 'type': 'function', 'index': 0}]}]}

>>>>>>> USING AUTO REPLY...
```

Query URL to arXiv

```
>>>>>>> EXECUTING FUNCTION search_file...
Searching via: http://export.arxiv.org/api/query?search_query=(cat:"hep-ex")+AND+(ti:"3770"+AND+ti:"pi")+AND+((ti:"3770"+OR+ti:"3770"+OR+ti:"*3770*")+AND+(ti:"pi"+OR+ti:"pi"+OR+ti:"*pi*"))+ANDNOT+((ti:"3686"+OR+ti:"3686"+OR+ti:"*3686*")+AND+(ti:"4660"+OR+ti:"4660"+OR+ti:"*4660*")+AND+(ti:"gluon"+OR+ti:"ball"+OR+ti:"gluon"+OR+ti:"ball"))&start=0&max_results=2&sortBy=submittedDate&sortOrder=descending
```

Query results (title, abstract, authors ...)

```
res = [0, {"0": {"Title": "Extracting strong phase and CP violation in  $D^0$  decays by using quantum correlations in  $B^0 \rightarrow D^0 \overline{D}^0$  correlations in  $(V_{1V_2})(K \rightarrow \pi \pi)$  and  $B^0 \rightarrow D^0 \overline{D}^0$  correlations in  $(V_{1V_2})(V_{3V_4})$ ", "First Author": "Hai-Bo Li", "Published date": "2010-10-08", "URL": "http://arxiv.org/abs/1010.1687v1", "Abstract": "We exploit the angular and quantum correlations in the  $D^0 \rightarrow \pi^+ \pi^-$  pairs produced through the decay of the  $\psi(3770)$  resonance in a charm factory to investigate CP-violation in two different ways. We consider the case of  $\psi(3770) \rightarrow D^0 \overline{D}^0 \rightarrow (V_{1V_2})(K \rightarrow \pi \pi)$  decays, which provide a new way to measure the strong phase difference  $\Delta\delta$  between Cabibbo-favored and doubly-Cabibbo suppressed  $D^0$  decays required in the determination of the CKM angle  $\gamma$ . We also build CP
```

Summary

- **AI era is coming**
 - It will not replace you but will help you to work more efficiently
- First AI agents for HEP - **Dr. Sai**
 - It can help people on data analysis (good accuracy for small task)
 - Many on-going studies to improve accuracy especially for complex task
 - Personpowers are more than welcome
- Next: more data, multi-modal, e.g. slides on indico, experts' feed back
 - **Please use it for your analysis or personal projects**
 - **We want and need your expertise !**
- Similar projects from other experiments:
 - LHC: [AccGPT](#) (LLaMa), AI assistant for accelerator control
 - ATLAS: [chATLAS](#)(GPT), AI assistant for internal navigator, same with one component of Dr. Sai
 - Outreach/education: [outreach assistant](#) (GPT), train people outside of HEP to analyze ATLAS open-data, same with one component of Dr. Sai

**We are ahead of them but still
have lots of works to be done !**

What can Dr. Sai do for STCF

- STCF vs BESIII
 - Almost the same except for detector/software
- Dr. Sai is designed for multiple usage
 - Physics analysis is benchmark
 - Easily transfer from BESIII to STCF
 - Many other potential application
 - Software upgrade/development
 - Internal navigator
 - CDR/TDR preparation
 - ...

back-up

Example: fitting

Prompt:

"I want to do two things:

1. generate a fitting script to fit a Gaussian distribution from the output file 'fake_data.root'. There is a TH1F name 'h_gaus' in the file.
2. use Tester to execute this code."

Example: BESIII assistant

Prompt:

"what is the path of inclusive MC sample produced at center-of-mass energy = 3.773 GeV."

what is the path of inclusive MC sample produced at center-of-mass energy = 3.773 GeV

Attention: auto-selected speaker 'Charm' is ready to handle the current task:
'what is the path of inclusive MC sample produced at center-of-mass energy = 3.773 GeV'.

Please give feedback if you have any comments (in 10 seconds).
1.'r' - Revise.
2.'c' or None - Continue.
3.'s' - Stop. ●

[ST]Ask human input for 10s[END]

No user input ●

To determine the path of the inclusive MC sample produced at a center-of-mass energy of 3.773 GeV, we need to identify the relevant information from the provided cases.

From the given cases, Case 3 provides specific information about the inclusive MC samples at 3.773 GeV:

```
Raw code
Sample Type
Sample Type,Ecm (GeV),RUNS,Luminosity(pb^-1) of DATA (not MC)
```

system tools

Please refer to Pan's talk for more details

```
Sample Type
Sample Type,Ecm (GeV),RUNS,Luminosity(pb^-1) of DATA (not MC)
on-ψ(3770) (2010),3.773,11414-13988,2931.8+-0.2+-13.8 *
on-ψ(3770) (2010),3.773,14395-14604,2931.8+-0.2+-13.8 *
on-ψ(3770) (2011),3.773,20448-23454,2931.8+-0.2+-13.8 *
```

Additionally, Case 4 provides the location of the dst files for MC samples:

```
Raw code
Location of dst Files
Directories for the MC samples can be found at:
BESIII DOC-228 http://docbes3.ihep.ac.cn/cgi-bin/DocDB/ShowDocument?docid=228
* Scripts were copied from
/panfs/panfs.ihep.ac.cn/home/data/bes3gen/workarea6XX/TestRelease/TestRelease-00-00-XX/psipp/
/ihepbatch/bes/ambrose/workarea-662/workarea/Reconstruction/QCMCFilterAlg/QCMCFilterAlg-00-00-01/share/jobOptions_QCMCFilter_makedst.txt
/besfs/offline/data/661-1/psipp/mc/4040/JobOptions/
```

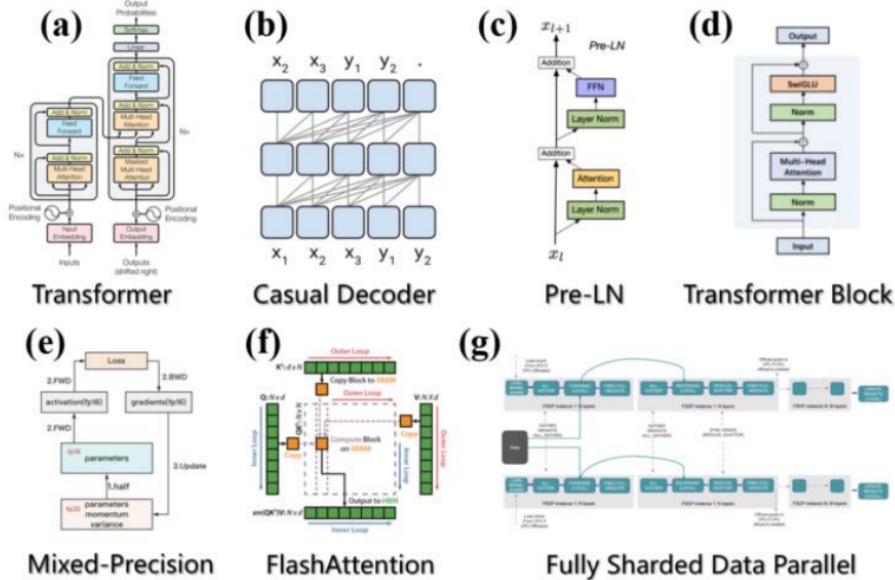
Combining this information, the path for the inclusive MC sample produced at a center-of-mass energy of 3.773 GeV can be inferred as follows:

```
Raw code
/besfs/offline/data/661-1/psipp/mc/3770/JobOptions/
```

This path is derived from the general structure provided in Case 4, adapted for the specific energy level of 3.773 GeV mentioned in Case 3.

Xiwu 2.0: Training techniques

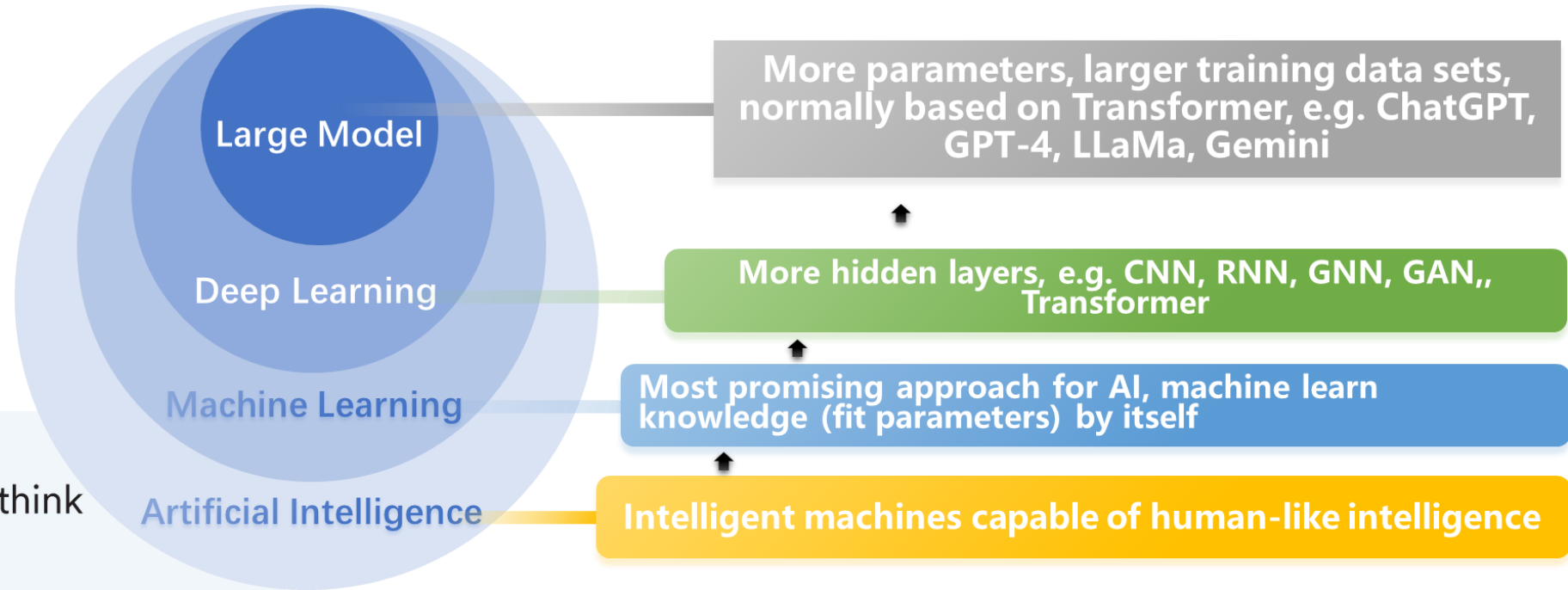
Frequently used techniques in LLM training



Reduce computational cost and training time, while improving model's performance

- **Mixed-precision Training**
reduce memory usage and improve computational throughput without sacrificing model accuracy
- **FlashAttention**
enhance floating-point operations per second (Flops)
- **Fully Sharded Data Parallelism (FSDP)**
parallel training across multiple devices
- **CPU-Offload**
offload certain computations from GPU to the CPU
- **Torchrun Parallelism**
Training on single node with Multi-GPU
- **Bitsandbytes**
vector quantization and mixed-precision decomposition
- **Deepspeed**
utilize mixed precision, gradient checkpointing, and offloading memory to the host CPU
- **ZeRO Series (ZeRO-2, ZeRO-3, ZeRO-offload)**
memory optimization strategies, such as activation checkpointing and optimizer state partitioning
- **Low-Rank Adaptive (LoRA)**
Approximate the weight matrices with low-rank factors while preserving the model's representational capacity

Machine learning and AI



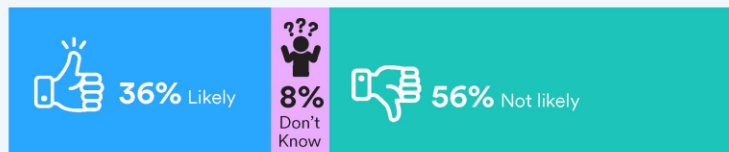
Global opinions: How people think AI will affect jobs, 2023

Source: Ipsos, 2023 | Chart: 2024 AI Index report

AI will change how you do your current job in the next 5 years



AI will replace your current job in the next 5 years

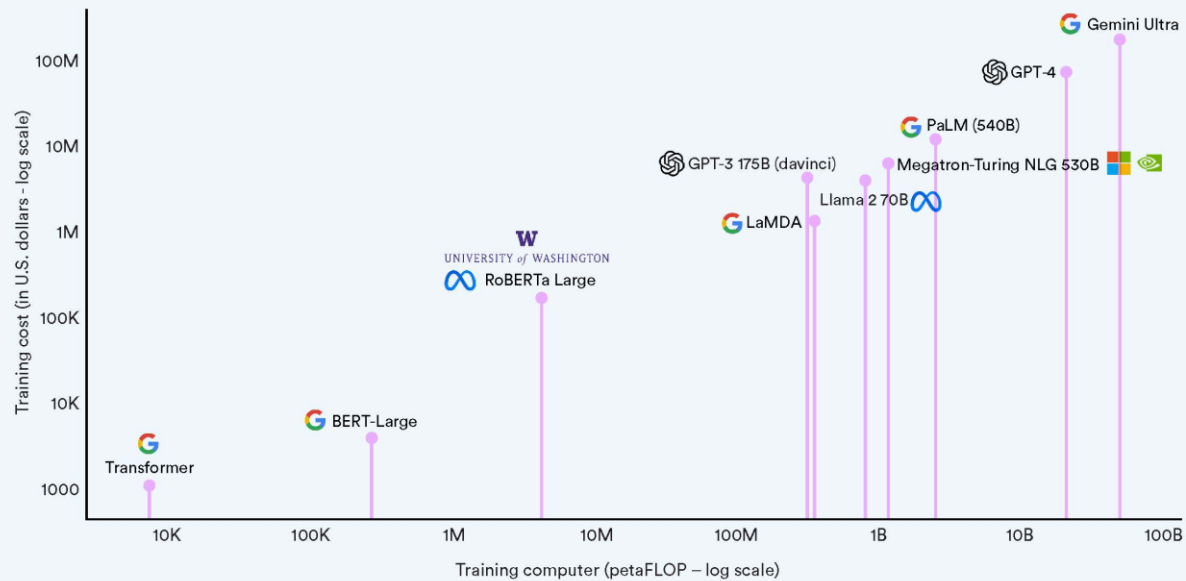


2024 AI index report

the next industrial revolution

Estimated training cost and compute of select AI models

Source: Epoch, 2023 | Chart: 2024 AI Index report



This is a C++ code for a class called `Gam4pikp` which is used to analyze data from the BaBar experiment. The class contains several methods for filtering and sorting data, as well as outputting results.

This is a C++ code for a data analysis algorithm called Gam4pikp. The algorithm is designed to analyze data from high-energy particle collisions and identify specific patterns of particles. The code appears to be a part of a larger program that is used to analyze data from the Large Hadron Collider (LHC) at CERN.

This is the implementation of a ROOT-based algorithm called `Gam4pikp` which is used to analyze particle physics data. The algorithm is designed to identify and reconstruct events containing four-pion (4π) and four-kaon ($4K$) resonances, and to perform various calculations and data analysis tasks.

This is a C++ code for a particle physics analysis tool, specifically a program that analyzes data from the Belle II experiment at the SuperKEKB collider. The code is designed to identify and reconstruct particles produced in high-energy collisions, and to study the properties of these particles.

Simple test: Dr. Sai vs GPT4

As expected, better performance in HEP

Q: 如何使用BOSS处理数据？

Answer from Dr. Sai

Answer from GPT-4

HaiChat Chat with AI Large Language Model v1.3 模型: GPT-4

如何使用BOSS处理数据

BOSS, 也就是Bit Oriented Synchronous Serial, 是一种数据通信协议, 但它并不直接用于数据处理。如果你指的是某种特定的数据处理软件或系统, 例如“BOSS系统”, 那可能需要更具体的信息来提供一个准确的答案。如果你是指一种名为“BOSS”的系统或软件, 一般来说, 使用这类系统处理数据的步骤可能包括以下几个步骤:

1. 数据输入: 首先, 你需要将数据输入到BOSS系统中。这可能涉及到上传文件, 或者通过API将数据导入系统。
2. 数据处理: 在数据被输入到系统之后, 你可以使用BOSS系统提供的工具和功能进行数据处理。这可能包括排序、过滤、聚合等操作。
3. 数据分析: 处理后的数据可以用于生成报告或进行进一步的分析。
4. 数据输出: 最后, 你可以将处理后的数据导出, 或者在BOSS系统中直接查看和分享。

不过, 具体的使用方法会根据BOSS系统的具体功能和你要处理的数据类型而变化。如果你能提供更多关于你正在使用的BOSS系统的信息, 我可能能提供更具体的帮助。

```

43 question = "如何使用BOSS处理数据"
44 # question = "给出一个BOSS上进行模拟的例子"
45 db_names = ["particle_physics_basic_concepts"]
46 print(f"Question: {question}")
47 print(f"Answer : ", end="")
48 answer = request_model(prompt=question, db_names=db_names,
49
50

```

问题 输出 终端 2

llm/chat_in_http.py
['hepai/chat', 'hepai/knowledge-DAO', 'openai/gpt-4', 'openai/gpt-3.5-turbo', 'meta/segment_anything_model', 'hepai/be-s-student-001']
Question: 如何使用BOSS处理数据
Answer : 要使用BOSS处理数据, 首先需要安装BOSS软件包并设置环境变量。然后, 按照以下步骤进行数据处理:

1. 模拟: 使用BOSS提供的模拟程序生成模拟数据。输入文件是一个配置文件, 输出文件是一个.rtraw文件。
2. 重建: 使用BOSS提供的重建程序对模拟数据进行重建。输入文件是.rtraw文件, 输出文件是一个.dst文件。
3. 分析: 使用BOSS提供的分析程序对重建数据进行分析。输入文件是.dst文件, 输出文件是一个.root文件。

每一步的输出文件格式如下:

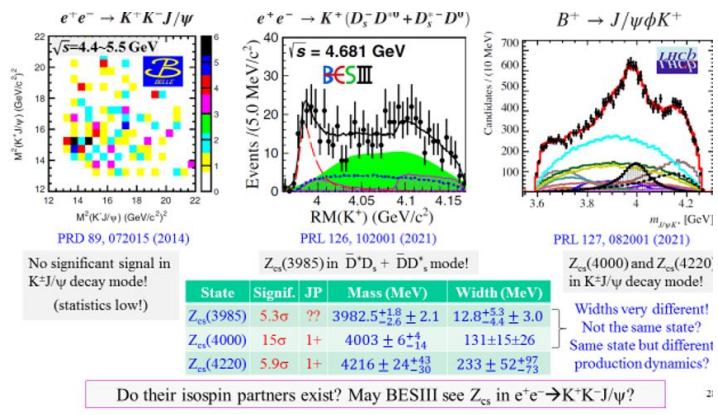
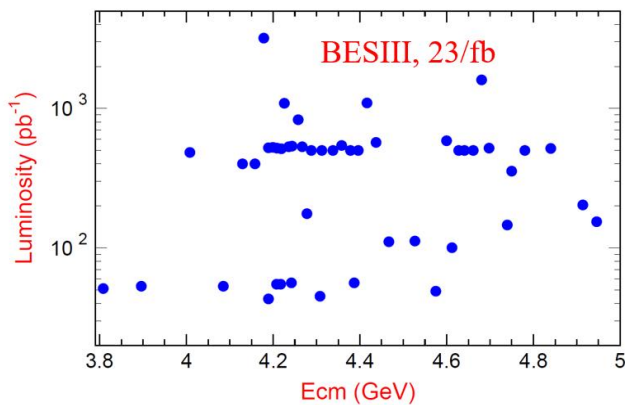
- 模拟: .rtraw文件是模拟数据的原始二进制文件。
- 重建: .dst文件是重建数据的二进制文件, 包含了事件的重建信息。
- 分析: .root文件是ROOT格式的文件, 其中包含了重建数据的分析结果, 可以使用ROOT软件包进行进一步的分析和可视化。

请注意, 以上是一般的BOSS数据处理流程和输出文件格式, 具体的流程和文件格式可能会根据实际需求和使用的BOSS版本而有所不同。建议参考BOSS的官方文档和示例程序以获取更详细的信息和指导。

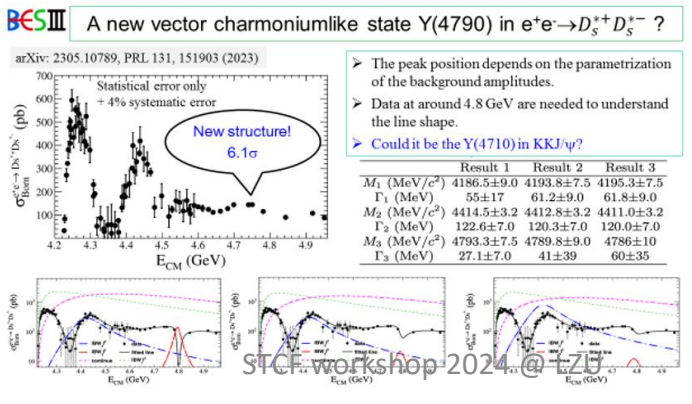
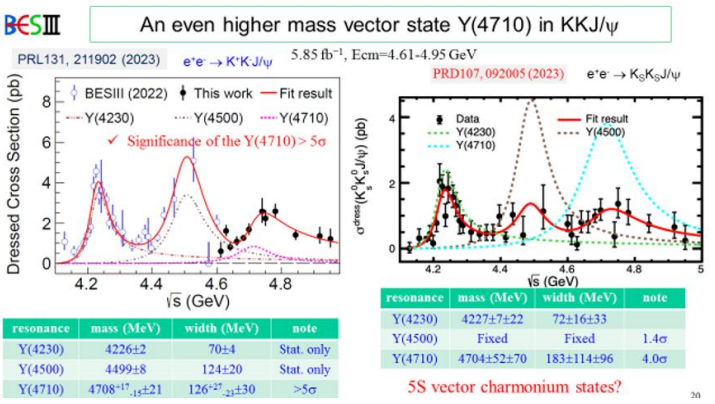
(langchain) → Re:Zc3900 git:(dev) X

Status and prospects for Dr. Sai

- XYZ studies at BESIII:
 - Many (intermediate) resonants, many decay channels -> interferences
 - Difficult to retrieve useful information correctly

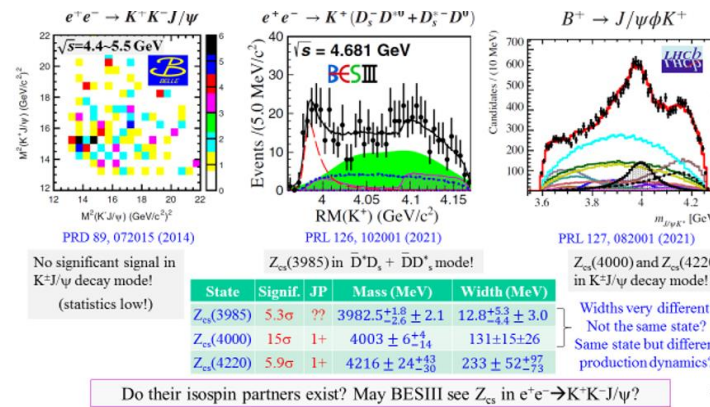
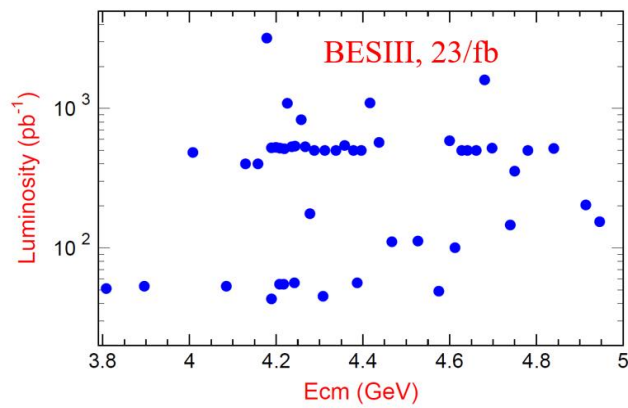


How to understand them? what is the physics behind them ?



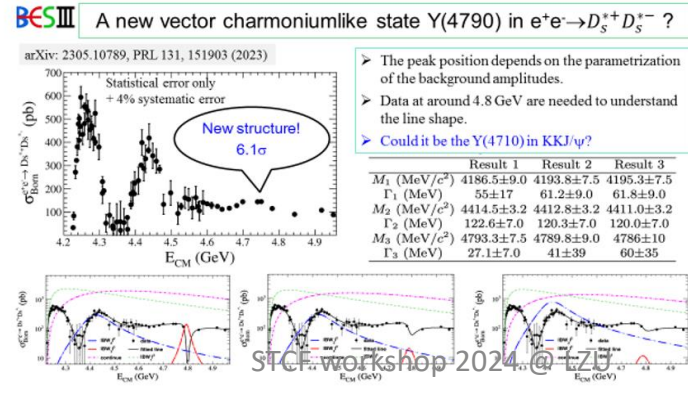
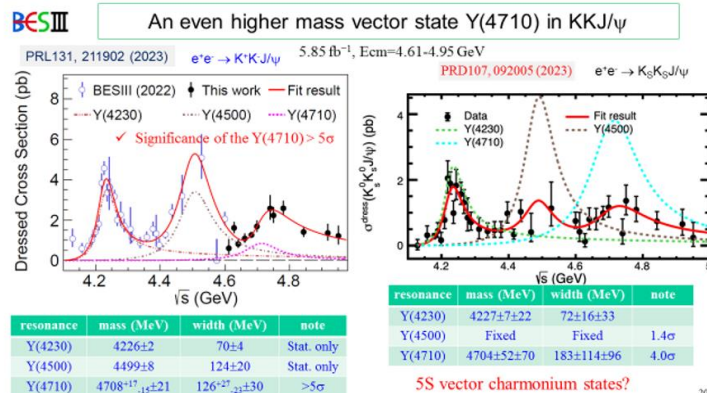
Status and prospects for Dr. Sai

- XYZ studies at BESIII:
 - Many (intermediate) resonants, many decay channels -> interferences
 - Difficult to retrieve useful information correctly



How to understand them? what is the physics behind them ?

As a experimental people, I don't know.



But the cross section measurements of ALL channels should be one necessary condition.

how many papers published from BESIII



The number of papers published by the BESIII collaboration can vary over time as new research is conducted and published. As of my knowledge cutoff date in October 2023, I don't have the exact count of papers published by BESIII.

To get the most up-to-date and accurate information, you can:

1. **Visit the BESIII Collaboration Website:** The official website of the BESIII collaboration may have a list of their publications or a link to their recent papers.
2. **Check Scientific Databases:** Use databases like the INSPIRE HEP (High Energy Physics) database, the arXiv preprint service, or journal publishers' websites to search for papers published by the BESIII collaboration.
3. **Contact the Collaboration Directly:** You can reach out to the BESIII collaboration or one of its representatives for detailed information on their publications.

These methods should provide the latest count and specifics of the academic papers published by the BESIII collaboration.

tion has published over 1,200 papers in peer-reviewed journals. This number is constantly growing as new data is

iment at the Beijing Electron Positron Collider (BEPC) in studying the properties of exotic hadrons and other aspects. The collaboration is one of the largest in the world, with over 1,000 institutions in 20 countries.

cover a wide range of topics, including the properties of states, the search for new physics beyond the Standard Model, and c reactions and processes.

Summary:

- **Production Channel:**

- e^+e^- collisions with ISR: ($e^+ e^- \rightarrow \gamma \pi^\pm X(3872)$)

- **Decay Channels:**

- Predominantly hadronic decays like ($X^\pm(3872) \rightarrow \pi^\pm J/\psi$)
- Radiative decays like ($X^\pm(3872) \rightarrow \gamma \pi^\pm J/\psi$)
- Multi-body decays where the final state includes charmonium.

By focusing on these channels, you can efficiently search for the charged partner of $X(3872)$ at BESIII.