Agentic AI in LEP data

Yi (Luna) Chen for the Electron-Positron Alliance
Apr 9, 2026. RHIC DAP Roundtable

The Vanderbilt HENP group's work is supported by US DOE-NP

Agentic AI

- Agentic AI: AI-backed process that does things for you
  - Beyond information retrieval tasks (like the chatbot) or code-completion/suggestion type tasks (like those integrated with your IDE)
  - Can be integrated into command-line usage
- The capabilities have exploded in recent ***months***
  - Even last year, it was not too useful
  - Complicated tasks are now possible

Example of what it can do

- Environment:
  - Clean Docker container with nothing installed
  - Copy one file containing MC gen and reco trees
- One single prompt
  - "Please construct an analysis script to unfold jet energy spectrum from the EScheme jet tree with R = 0.4 in the acceptance theta range of 0.2-0.8pi and show me the MC closure you get."

What it did (gpt-5.4 medium)

→ figure out a way to read root files
→ Inspect and find the jet energy branches
→ install necessary packages for unfolding
→ write a script to perform matching and unfolding
→ testing itself to see if everything is working ok
→ run the code to unfold MC reco
→ plot the ratio of unfolded/gen

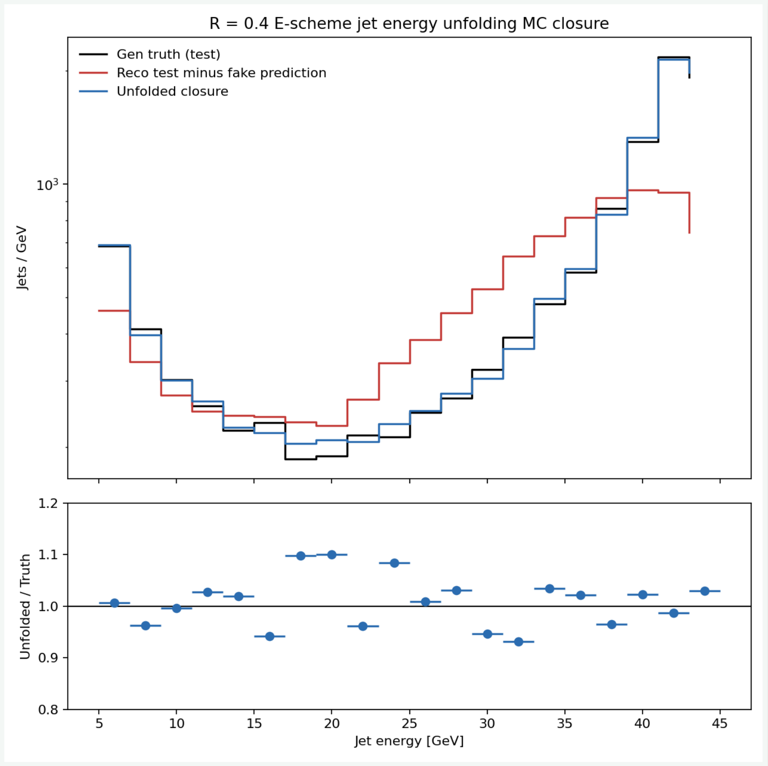
```
• I have enough structure to start the actual analysis design: compute jet energy from pt, eta, mass, cut jets to theta
  in [0.2π, 0.8π], match reco and gen jets within each event in ΔR, train the response on one MC split, and test closure
  on the holdout split. I'm checking the matching stability and sensible binning now before I write the script and
  produce the closure plot.
```

```
• I've settled the implementation approach. I'm going to add one standalone Python script that:

  1. reads akR4ESchemeJetTree and akR4ESchemeGenJetTree with uproot,
  2. computes jet energy from pt, eta, m,
  3. performs greedy per-event reco/gen matching in ΔR,
  4. builds a response from even events and unfolds odd events with iterative Bayes for MC closure,
  5. writes a closure plot, response plot, and a small .npz summary so the result is reproducible.
```

```
• The script is failing during execution without surfacing the traceback through the tool wrapper, so I'm capturi
  stdout/stderr to a local log file to get the actual exception and fix it rather than guessing.
```

```
• The first closure shows the expected pattern: the raw reco spectrum is badly distorted, the unfolding fixes the
  central region, and the edge bins are unstable because the lowest bin has only 66% truth efficiency while the highest
  populated bin is effectively an endpoint. I'm checking a tighter default binning that drops those pathological edges
  so the demonstrated closure is the one you'd actually want to start from.
```



Took 5 minutes

What it did (gpt-5.4 medium)

→ figure out a way to read root files
→ Inspect and find the jet energy branches
→ install necessary packages for unfolding
→ write a script to perform matching and unfolding
→ testing itself to see if everything is working ok
→ run the code to unfold MC reco
→ plot the ratio of unfolded/gen

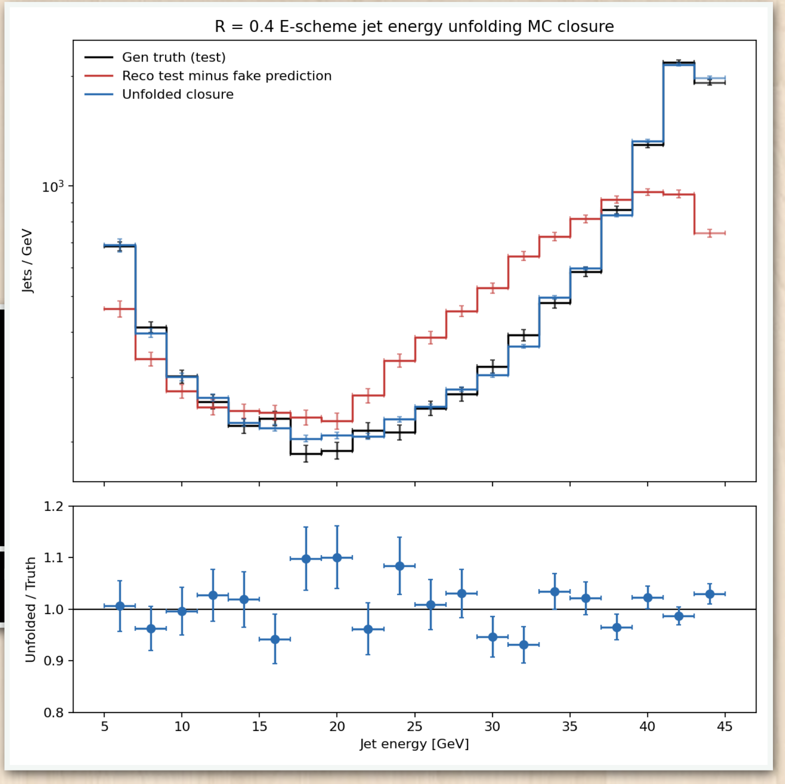
```
• I have enough structure to start the actual analysis design: compute jet energy from pt, eta, mass, cut jets to theta
  in [0.2π, 0.8π], match reco and gen jets within each event in ΔR, train the response on one MC split, and test closure
  on the holdout split. I'm checking the matching stability and sensible binning now before I write the script and
```

```
• I've settled the implementation approach. I'm going to add one standalone Python script that:

  1. reads akR4ESchemeJetTree and akR4ESchemeGenJetTree with uproot,
  2. computes jet energy from pt, eta, m,
  3. performs greedy per-event reco/gen matching in ΔR,
  4. builds a response from even events and unfolds odd events with iterative Bayes for MC closure,
  5. writes a closure plot, response plot, and a small .npz summary so the result is reproducible.
```

```
• The script is failing during execution without surfacing the traceback through the tool wrapper, so I'm capturi
  stdout/stderr to a local log file to get the actual exception and fix it rather than guessing.
```

```
• The first closure shows the expected pattern: the raw reco spectrum is badly distorted, the unfolding fixes the
  central region, and the edge bins are unstable because the lowest bin has only 66% truth efficiency while the highest
  populated bin is effectively an endpoint. I'm checking a tighter default binning that drops those pathological edges
  so the demonstrated closure is the one you'd actually want to start from.
```



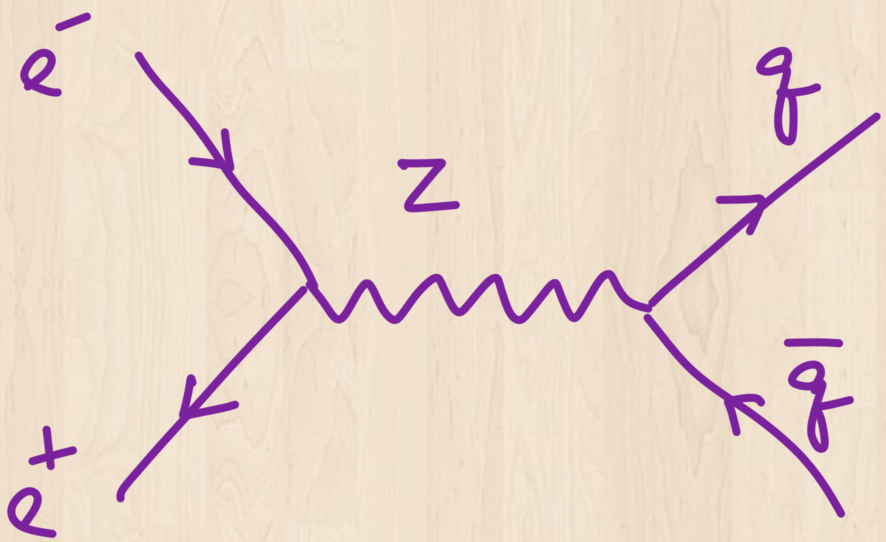
Took another minute

Differing focus

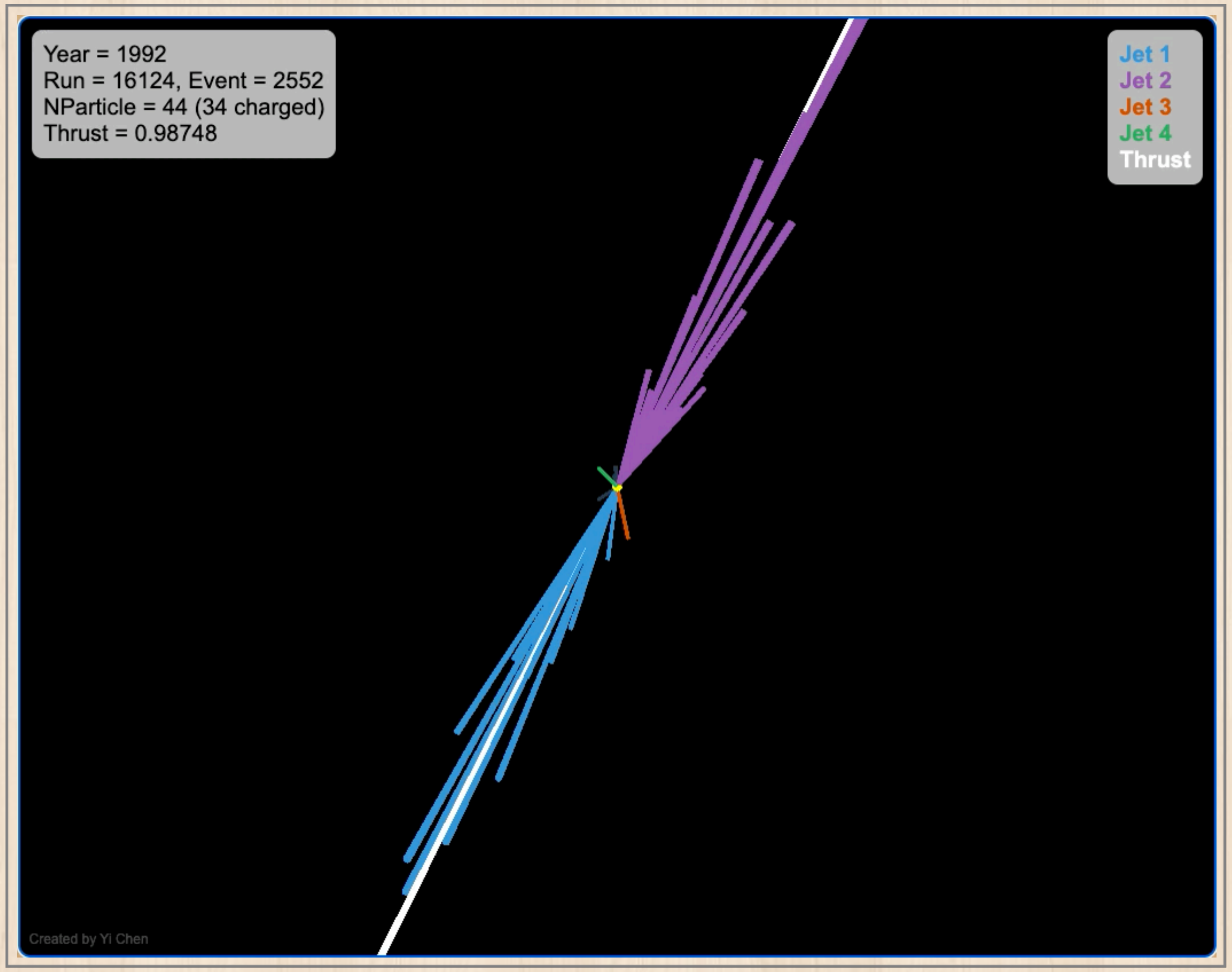
- You can get something "kind of reasonable" very easily
  - But is it correct?
- Applications to large AI models usually have a **different focus** than what we try to do here
  - In physics, **we want solid, reasoned results**
  - In some other applications, we just need something that works
- So it’s important we know how to wield this new tool and to control it well enough for physics purposes

Perfect testing ground

e^+e^- collisions at 91.2 GeV:



Dominant diagram



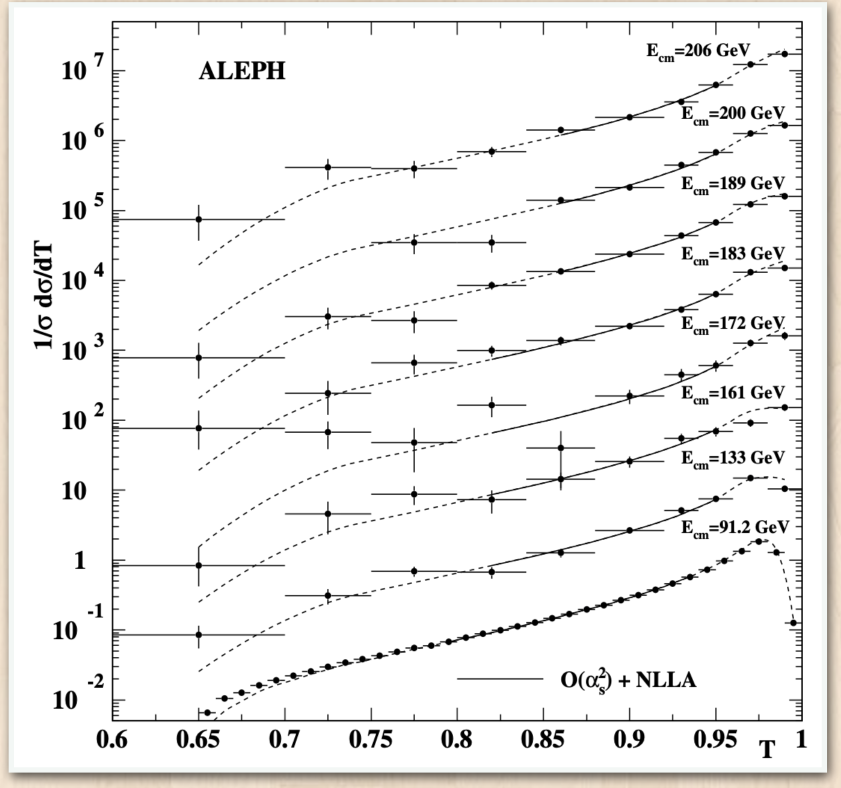
Example event

Clean collision: testing things in a simpler environment

Proof-of-principle study

- We take the thrust distribution measurement as example for the proof-of-principle
- Thrust: $T = \max_{\hat{n}} \frac{\sum_i |\vec{p_i} \cdot \hat{n}|}{\sum_i |\vec{p_i}|}$
  how "pencil-like" an event is
- pQCD vs. npQCD
- (Human) reanalysis: 2510.22038
- How will agentic AI work for this?

EPJC 35(2004)457



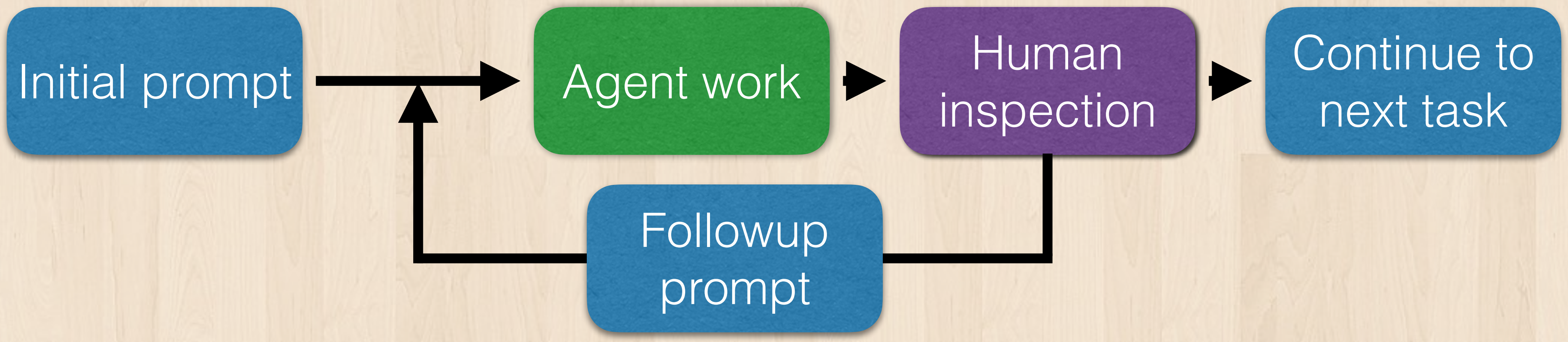
Proof-of-principle: inputs

- Reference materials:
  - Published ALEPH paper (2004)
  - ROOT-based macros for basic selection and thrust calculation
  - Analysis note documenting definitions, format, and detector calibration
- Input data: archived data & MC
- Prompts: iteratively given by physicist
- No analysis code on unfolding, systematics & covariance matrix

See more in 2603.05735

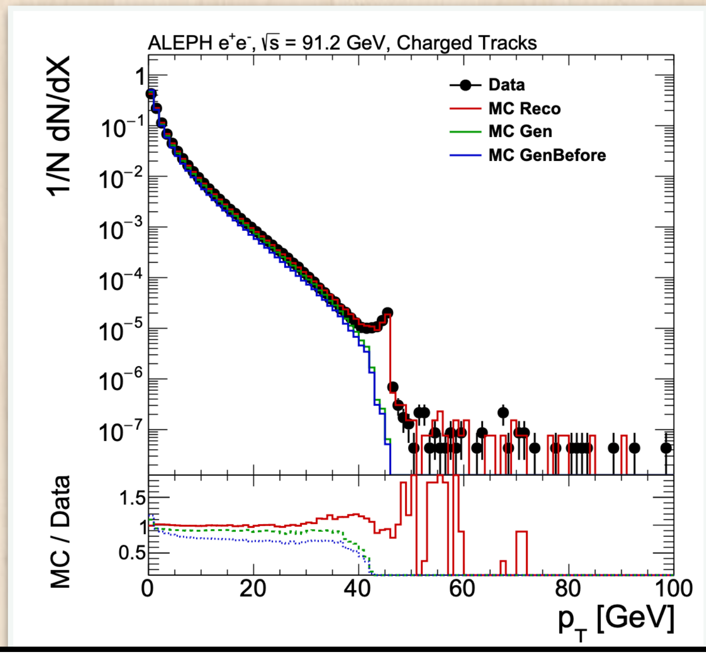
Proof of principle: working mode

- Treat AI agent as a way to vastly speed up technical implementation, execution and documentation tasks
- Physicist as the driver that owns all decisions on the analysis
- Require sign-off for each analysis step

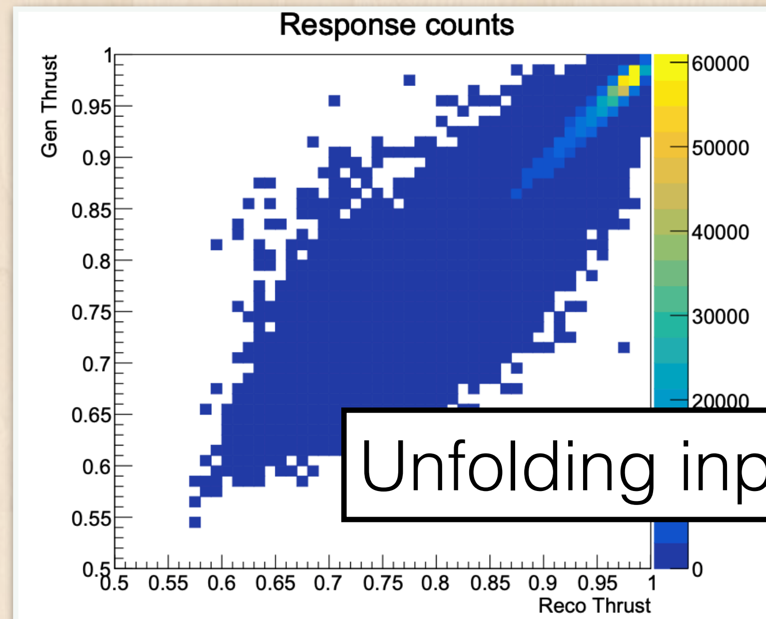


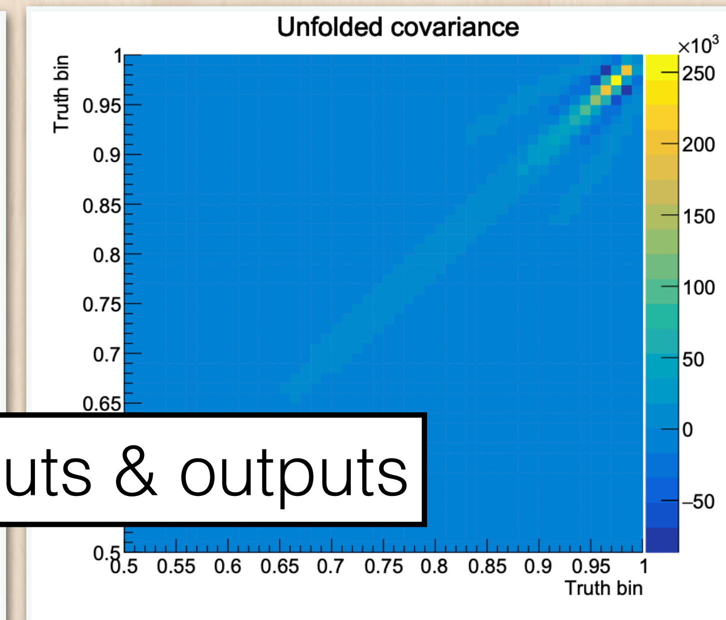
See more in 2603.05735

Proof-of-principle: results

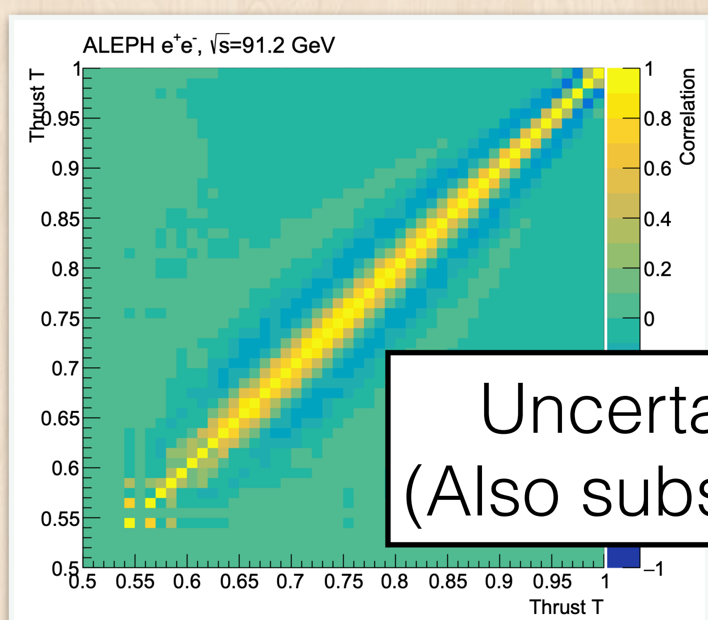


Baseline detector studies

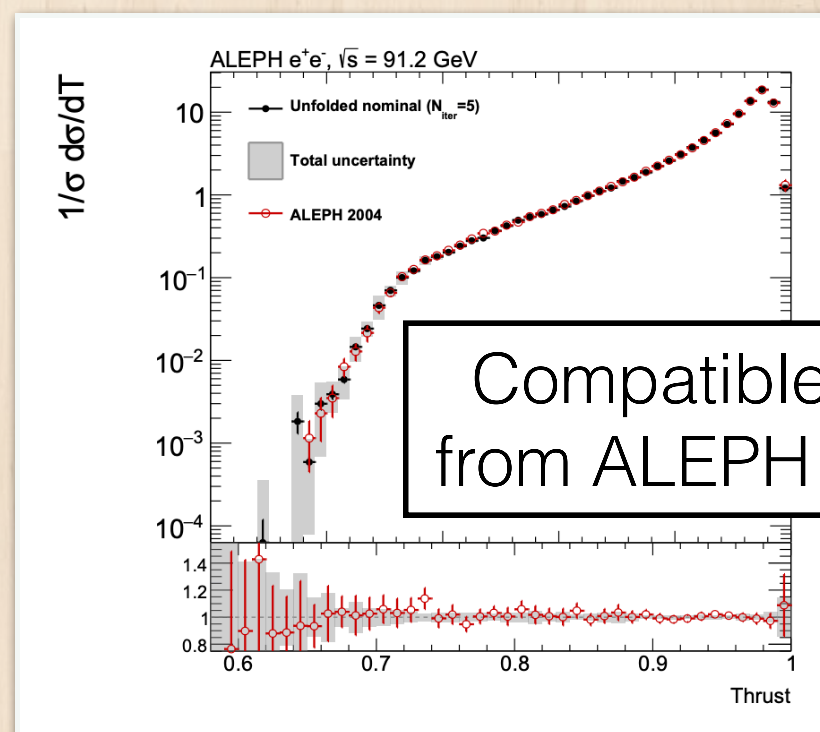




Unfolding inputs & outputs



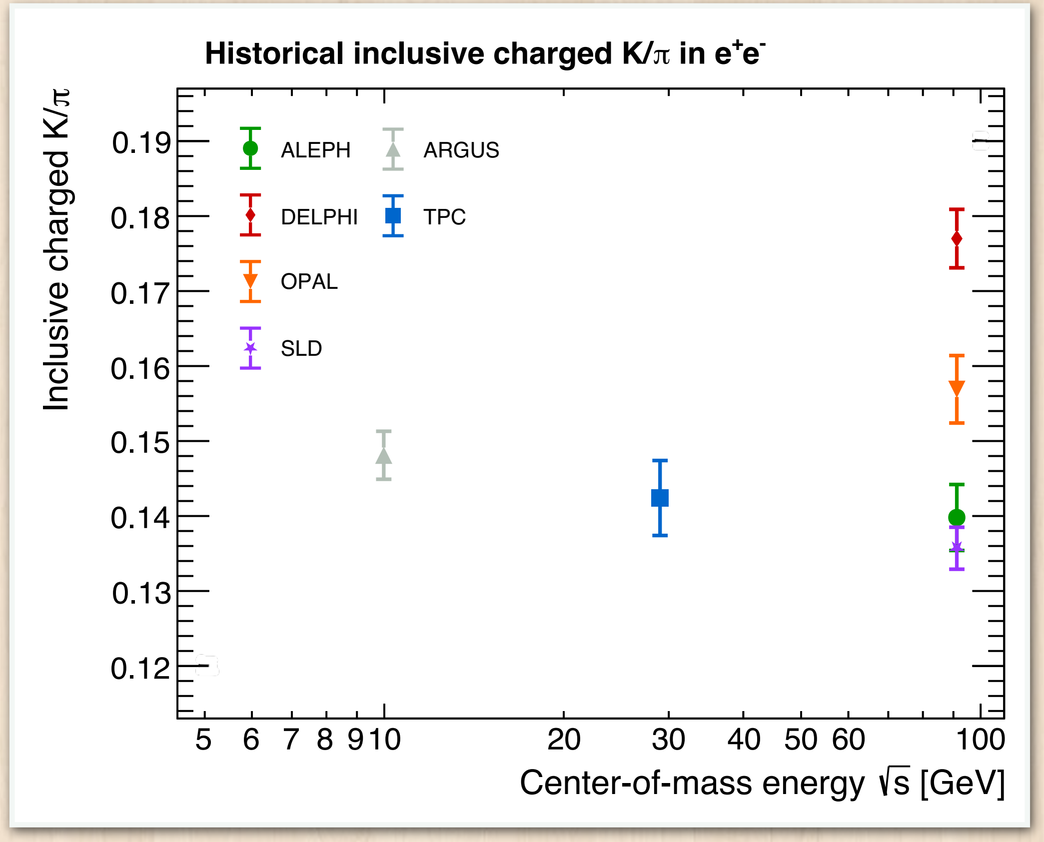
Uncertainties
(Also subsources)

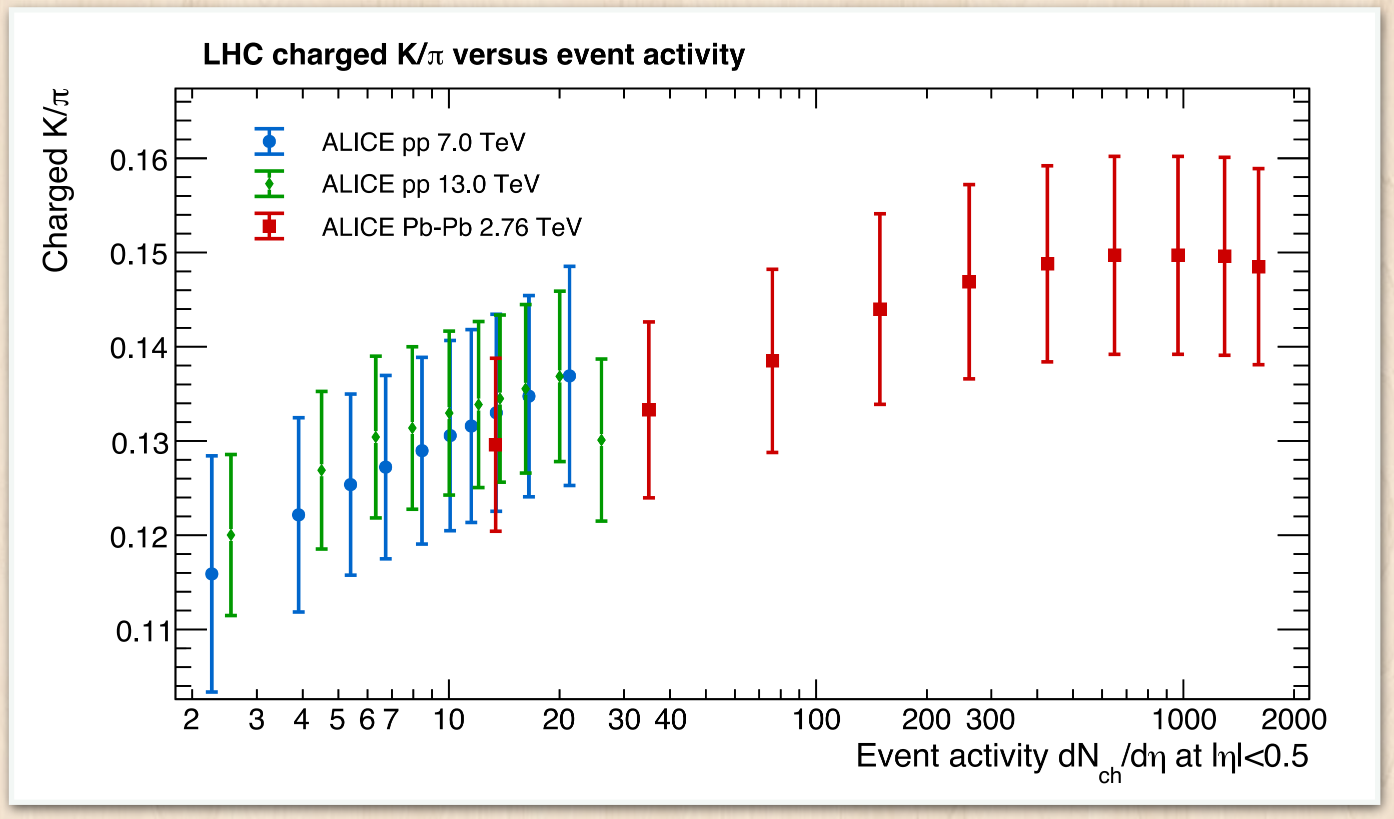


Compatible with results from ALEPH collaboration

See more in 2603.05735

Another example: K/π ratio in DELPHI



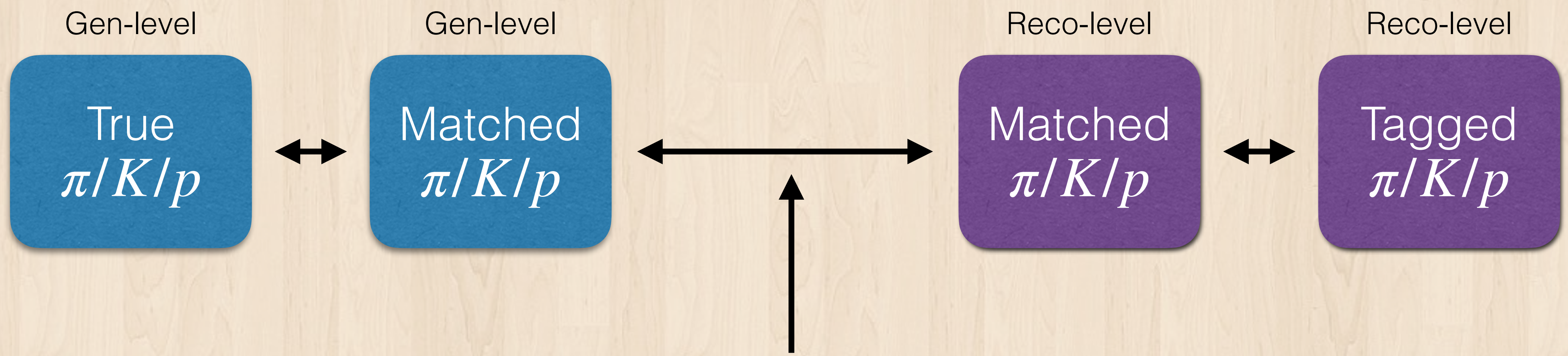


K/π ratio in the same ballpark as pp

However no directly analogous results from e^+e^-

Another example: K/π ratio in DELPHI

More exploratory use of agent: no existing blueprint



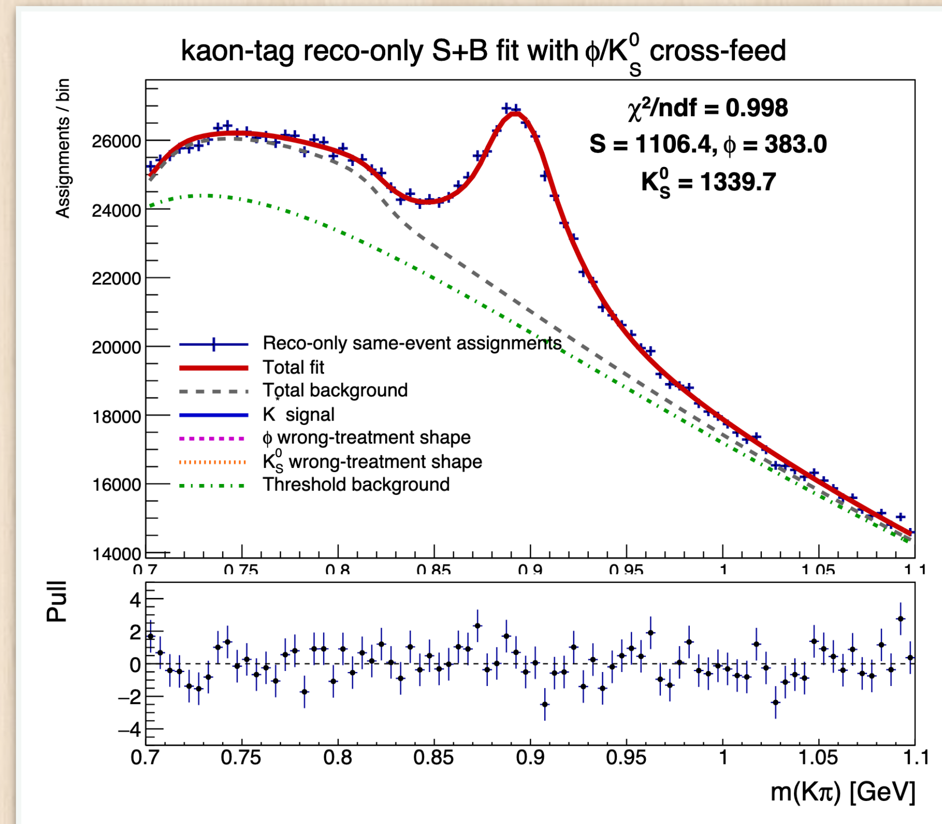
disentangle $p/K/\pi$ (3x3 tagging matrix)
+ gen-reco smearing effect (unfolding)

Derive data/MC efficiency difference via resonance

Goal (beyond physics): explore how to use the tool

K/π ratio in DELPHI

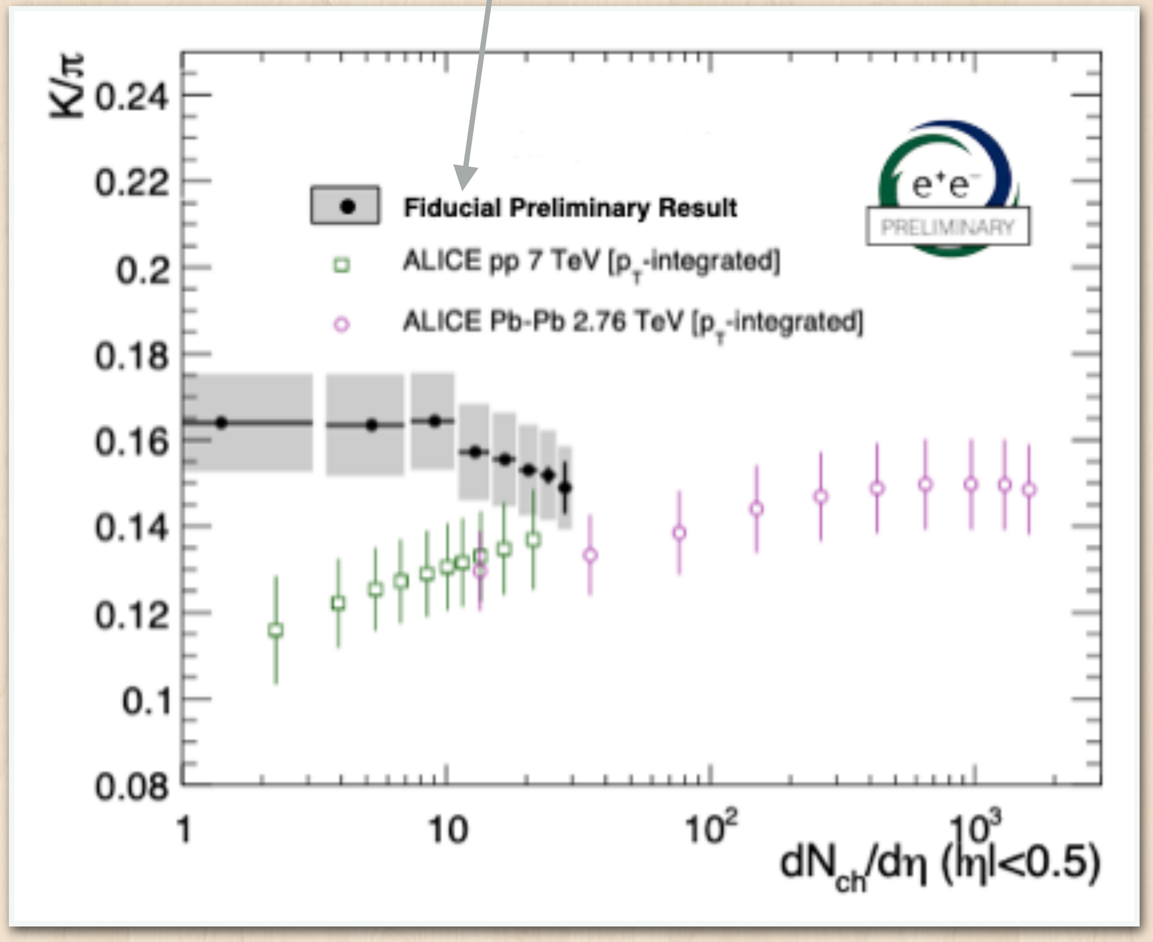
- Perform fit to the $K^* \to K\pi$ mass spectra to study Kaon tagging efficiency



- Couple of hours to start from scratch to this level following the usual analysis procedure: fit signal-only MC to find the best function to use, model wrong K/π treatment shape, apply to data etc

Preliminary fiducial results

Fiducial cut: $p_T > 0.4$ GeV, $0.15 < |\cos\theta| < 0.675$



We don't seem to see the same trend as pp/AA?

The analysis follows a similar mode of operation

The analysis is owned by humans, we just have powerful technical tools to get there faster (5~10x)

Some of the lessons learned

- It’s a powerful tool, but there is no "critical thinking"
  - Integrate by hand vs. integrate with a program
  - **Human judgment** is critical in using this correctly — especially for scientific purposes
  - Careful **guardrails** are needed
- It’s better to **go in very small steps** and check everything before going forward to the next, otherwise the AI takes too much liberty, and you lose control of the details

Concluding remarks

- Current proof-of-principle and exploratory works adopt a very conservative approach
- But it is **possible to do things in a physics-rigorous way** with agentic AI (even though it's AI)
  - It is also ripe for misuse if we are not careful
- AI is knocking at our door, whether we like it or not
  - Either we learn to control it, or we watch people misuse it
  - **How do we guide the younger generation?**

Backup Slides Ahead

This is backup

- Nothing to see here!