LLM Use in ePIC at Scale

Goal: facilitate interaction of LLMs with ePIC software and computing throughout!

Physics analysis:

“Create a two-pion invariant mass spectrum for the latest 10x100 production campaign of ePIC, using all available Q2 bins weighted by their cross section.”

Detector performance:

“Determine the average energy resolution for gammas in the EEEMCal as a function of the gamma energy, and fit the curve with a typical calorimetry resolution fit.”

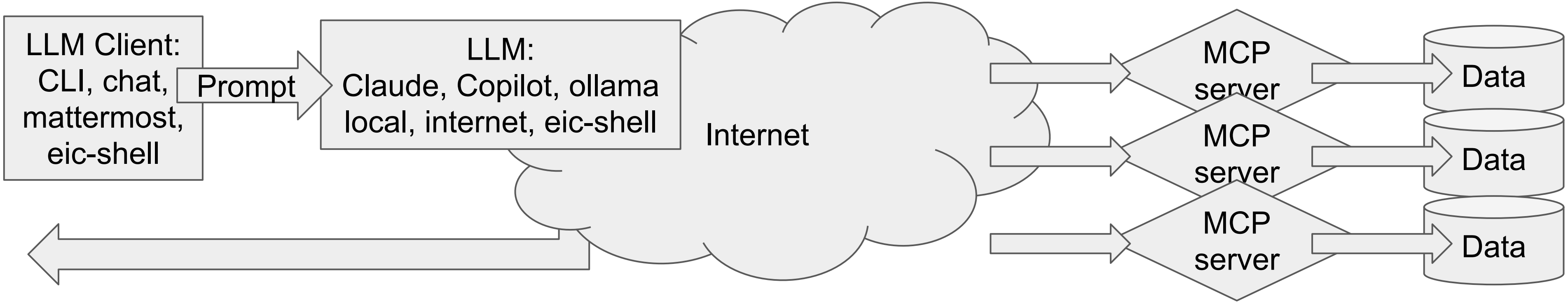
Detector geometry:

“Compare the geometry implementation of the SVT positive endcap disks with the latest values in the detector parameter tables or the CAD model from the project.”

Already reasonably available

ePIC specific interface needed

LLM Use in ePIC at Scale



Interaction of LLMs with specific data sources happens via MCP servers:

- Model Context Protocol: “API” layer for the LLM to get access to data sources
- e.g. GitHub MCP server provides access to repository details (commits, comments, CI runs and logs)

Providing access to ePIC specific data sources will benefit from **MCP server development or configuration/provisioning of existing MCP servers.**

LLM Use in ePIC at Scale

Existing MCP servers:

- GitHub MCP servers (default in Copilot)
- AskPanDA, XRootD MCP servers

Potential new MCP servers:

- Geometry Database MCP server (interfaces with latest geometry table)
- CAD model MCP server (interprets questions about CAD models/dimensions)
- Zenodo MCP server (access to documents*)

Issues for consideration:

- MetaMCP at mcp.epic-eic.org for easy config of all available MCP servers
- Development is straightforward and essentially just connecting APIs
- Safeguard access to ePIC only authenticated resources (with *)

LLM Use in ePIC at Scale

Physics analysis:

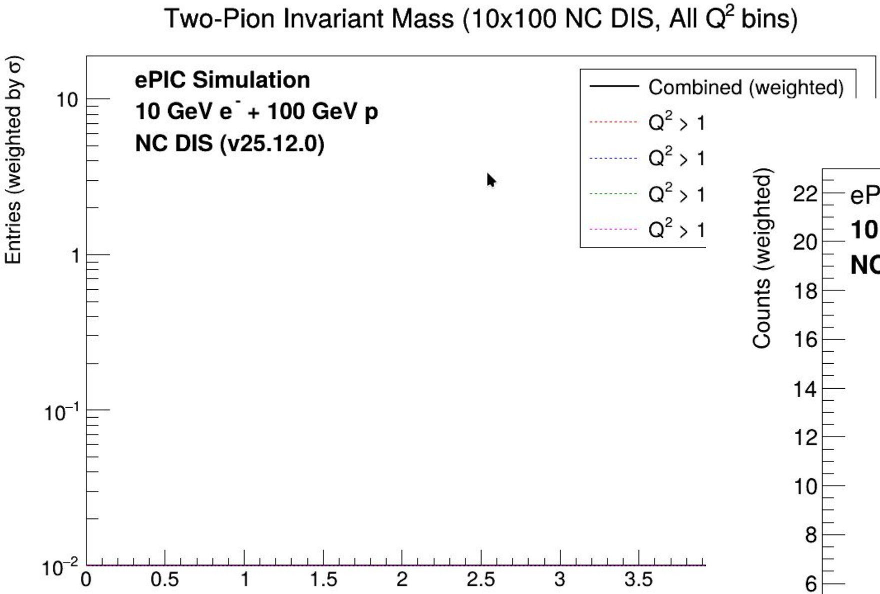
> “Create a two-pion invariant mass spectrum for the latest 10x100 production campaign of ePIC, using all available Q2 bins weighted by their cross section.”

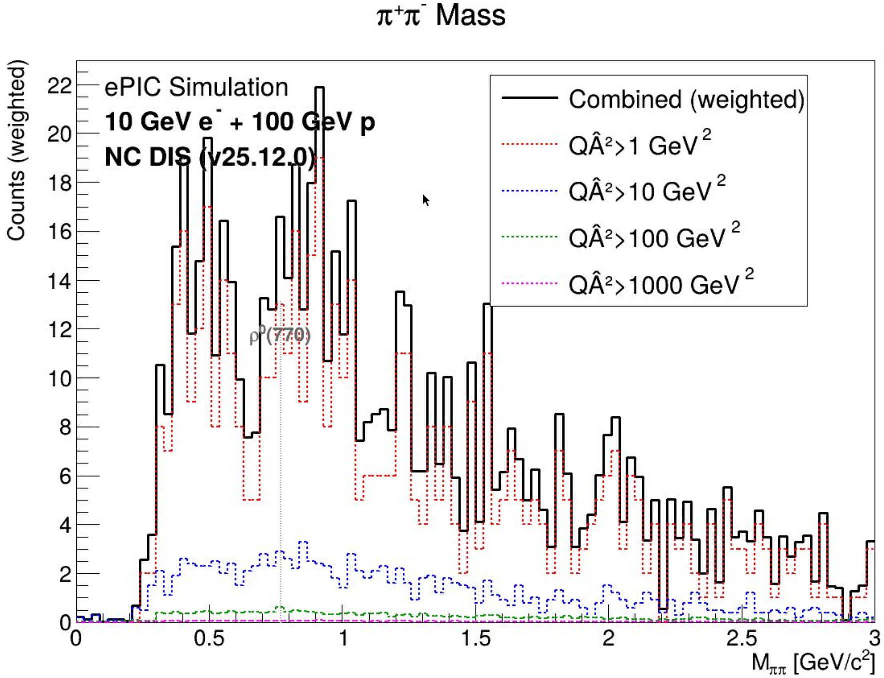
In context of EICrecon repository (outside eic-shell): (with minimal additional instructions)

- Uses `xrdfs` to query dtn-eic.jlab.org (XRootD MCP server was not connected)
- Creates python script to run over all relevant files
- Realizes too much data, restarts…
- Assumes weights proportional with Q2 instead of a using actual cross section (but I don’t know where to get that info either…)
- Claims to be done when image is empty…
  “The spectrum shows the expected ρ^0 meson structure at ~0.770 GeV/c² ($\rho^0 \rightarrow \pi^+\pi^-$ decay) plus continuum background from non-resonant pion production in deep inelastic scattering.”
- Finally repeats with only 200 events per energy setting

B

LLM Use in ePIC at Scale





LLM Use in ePIC at Scale

Detector performance:

"Determine the average energy resolution for gammas in the EEEMCal as a function of the gamma energy, and fit the curve with a typical calorimetry resolution fit."

In context of EICrecon repository (outside eic-shell): (with minimal additional instructions)

- Runs full `epic_craterlake` simulation with ddsim over various input energies
- Attempts to run `eicrecon` over data but got hung up on missing collections
- Recompiles all of `eicrecon`.

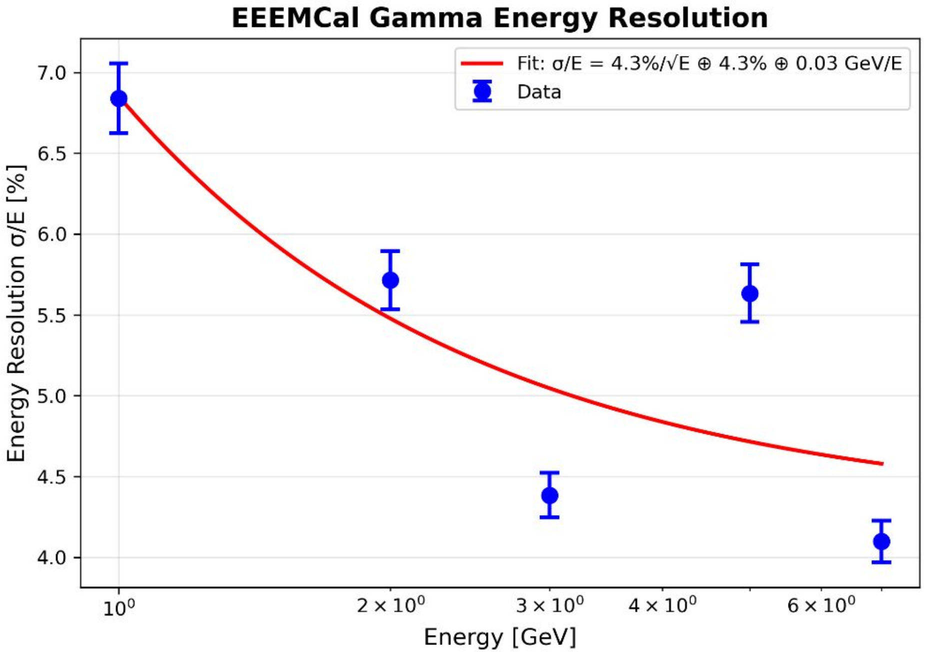
In context of EICrecon repository (inside eic-shell): (with minimal additional instructions)

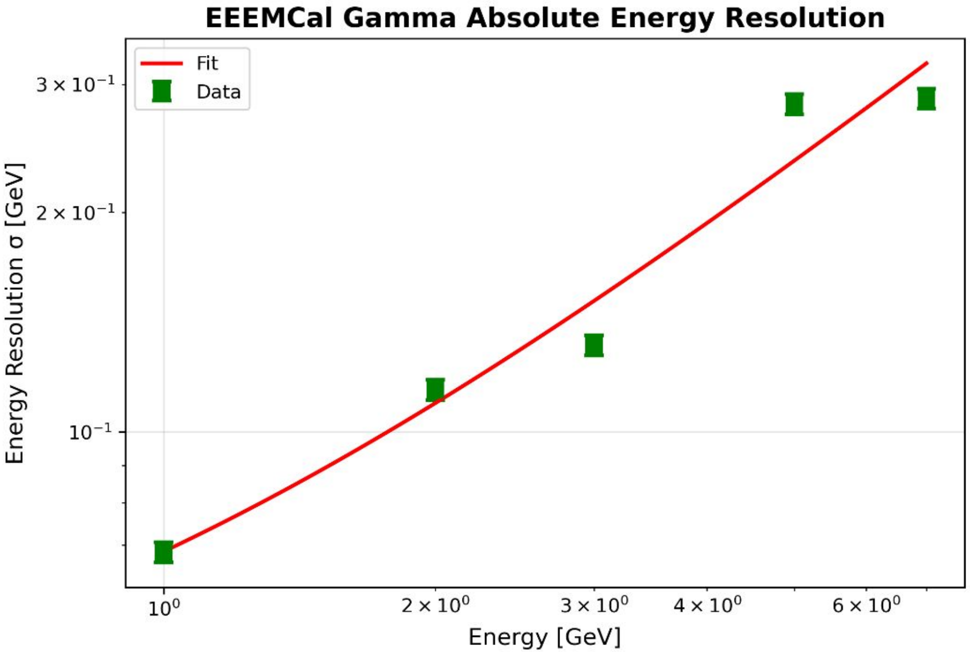
- Runs full `epic_craterlake` simulation with ddsim over various input energies
- Warns this will take a long time… switches to 100 events per data point (1, 2, 3, 5, 7 GeV)
- Incorrectly uses `-Pjana:nevents=100 -Pjana:nthreads=1` (consistency is key)

Ultimately $\sigma/E = (4.34\% \pm 7.89\%) / \sqrt{E[\mathrm{GeV}]} \oplus (4.25\% \pm 1.64\%) \oplus (0.032 \pm 0.094)\ \mathrm{GeV} / E[\mathrm{GeV}]$

LLM Use in ePIC at Scale

B+





LLM Use in ePIC at Scale

Fundamental concerns and potential pitfalls:

- Reproducibility and auditability of LLM results
    - Not just happy with the answer, but must be able to verify
        - Difference from code: tests as the definition of done
    - Direct LLM to produce reproducible and auditable results
- Equitable access to LLM ‘horsepower’
    - Integration in collaboration tools, like mattermost and eic-shell, rather than simply directing people to buy their own LLM API keys