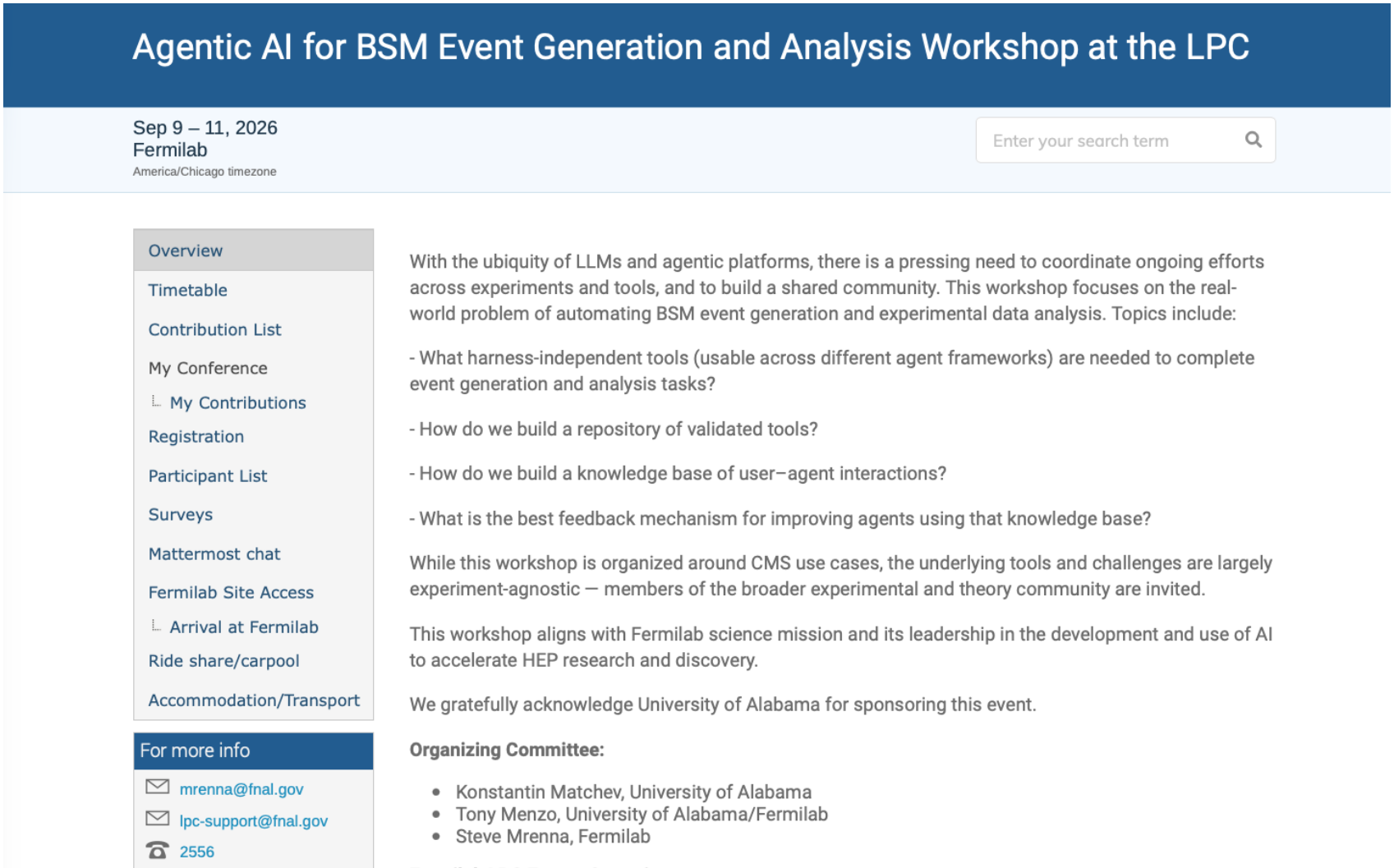
A Call for an AI Network for HEP(N) Theory

Stephen Mrenna
Fermilab

09-14-26

https://indico.fnal.gov/event/75639/



Agentic AI for BSM Event Generation and Analysis Workshop at the LPC

Sep 9 – 11, 2026
Fermilab
America/Chicago timezone

Enter your search term

- Overview
- Timetable
- Contribution List
- My Conference
  - My Contributions
- Registration
- Participant List
- Surveys
- Mattermost chat
- Fermilab Site Access
  - Arrival at Fermilab
- Ride share/carpool
- Accommodation/Transport

For more info

mrenna@fnal.gov
lpc-support@fnal.gov
2556

With the ubiquity of LLMs and agentic platforms, there is a pressing need to coordinate ongoing efforts across experiments and tools, and to build a shared community. This workshop focuses on the real-world problem of automating BSM event generation and experimental data analysis. Topics include:

- What harness-independent tools (usable across different agent frameworks) are needed to complete event generation and analysis tasks?

- How do we build a repository of validated tools?

- How do we build a knowledge base of user–agent interactions?

- What is the best feedback mechanism for improving agents using that knowledge base?

While this workshop is organized around CMS use cases, the underlying tools and challenges are largely experiment-agnostic — members of the broader experimental and theory community are invited.

This workshop aligns with Fermilab science mission and its leadership in the development and use of AI to accelerate HEP research and discovery.

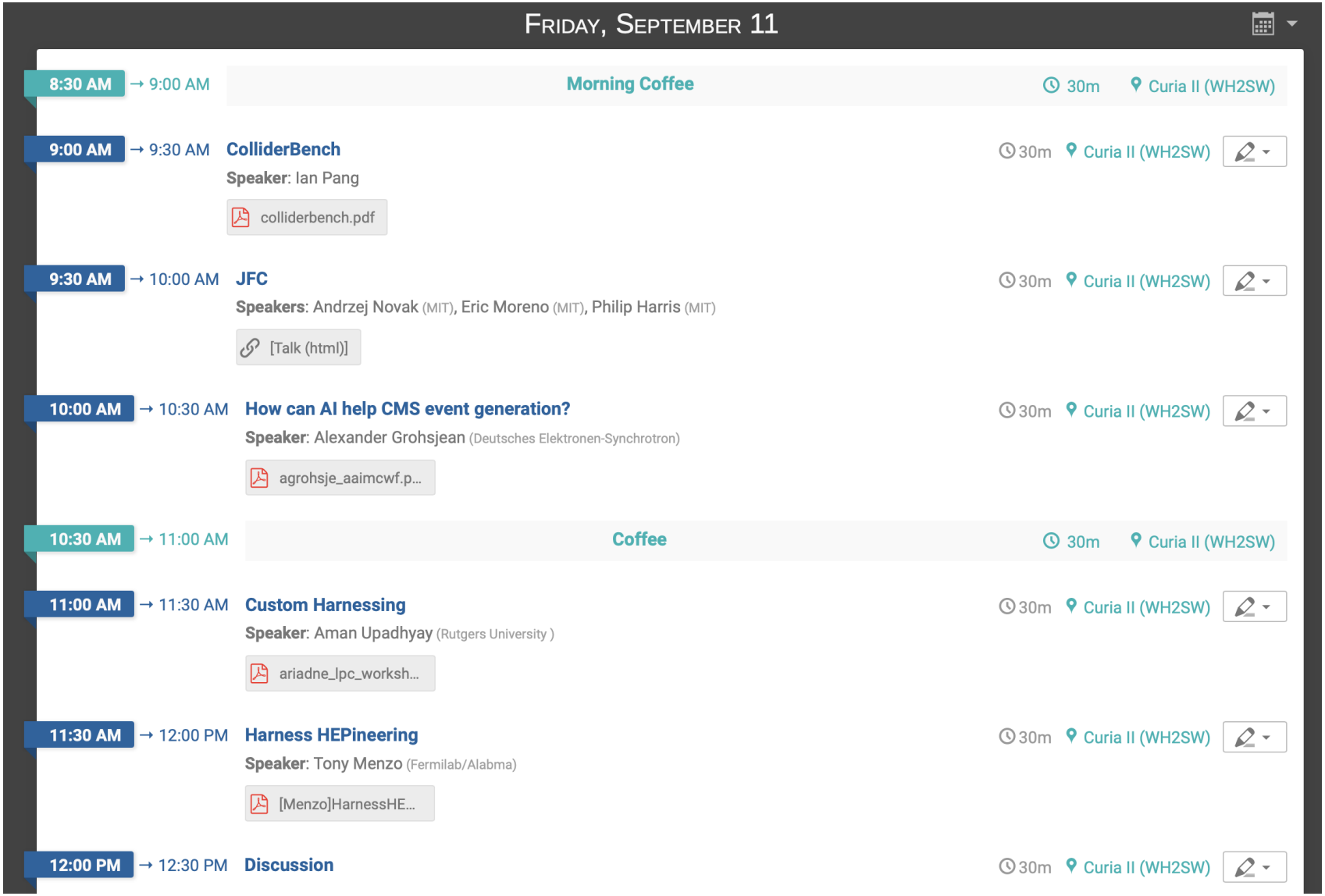
We gratefully acknowledge University of Alabama for sponsoring this event.

Organizing Committee:

- Konstantin Matchev, University of Alabama
- Tony Menzo, University of Alabama/Fermilab
- Steve Mrenna, Fermilab

64 registrants: most remote, but a fair number in person.
Summary of talk given here.

https://indico.fnal.gov/event/75639/



FRIDAY, SEPTEMBER 11

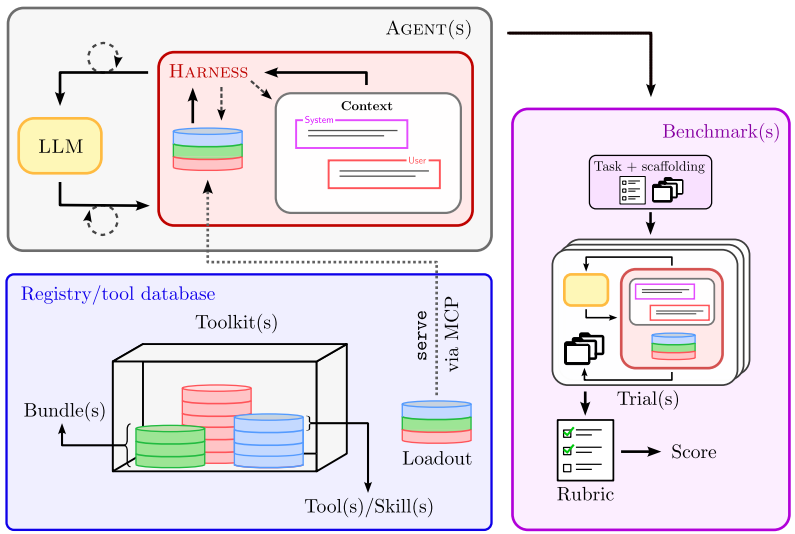
8:30 AM → 9:00 AM	Morning Coffee		30m	Curia II (WH2SW)
9:00 AM → 9:30 AM	ColliderBench Speaker: Ian Pang colliderbench.pdf		30m	Curia II (WH2SW)
9:30 AM → 10:00 AM	JFC Speakers: Andrzej Novak (MIT), Eric Moreno (MIT), Philip Harris (MIT) [Talk (html)]		30m	Curia II (WH2SW)
10:00 AM → 10:30 AM	How can AI help CMS event generation? Speaker: Alexander Grohsjean (Deutsches Elektronen-Synchrotron) agrohsje_aaimcwf.p...		30m	Curia II (WH2SW)
10:30 AM → 11:00 AM	Coffee		30m	Curia II (WH2SW)
11:00 AM → 11:30 AM	Custom Harnessing Speaker: Aman Upadhyay (Rutgers University) ariadne_lpc_worksh...		30m	Curia II (WH2SW)
11:30 AM → 12:00 PM	Harness HEPineering Speaker: Tony Menzo (Fermilab/Alabma) [Menzo]HarnessHE...		30m	Curia II (WH2SW)
12:00 PM → 12:30 PM	Discussion		30m	Curia II (WH2SW)

We are talking about many common topics!

https://indico.fnal.gov/event/75639/

Utilize and adapt existing infrastructure as much as possible!
General purpose coding harnesses + domain tools & skills = HEP agent

- **Adopt a standard protocol for interfacing tools with harnesses**
  - **Model Context Protocol (MCP)** has been widely adopted by all harnesses, let's not reinvent the wheel!
- **Community maintained registry/tool-skill database + serving**
  - Scalable – what do you do when you have $\mathcal{O}(1000)$ tools to choose from? Need easy curation of custom "loadouts"
- **Science-forward benchmarking framework**
  - Emphasis on reproducibility and full "6-sphere" resolution



Parallel developments need coordination!

The Call: An AI-oriented network for HEP(N) theorists

One Wing of an HEP AI Effort

- New AI tools are changing physics research
- The tools we need are not cheap.
- Science funding has moved to AI.
- Experimental HEP is already organizing.

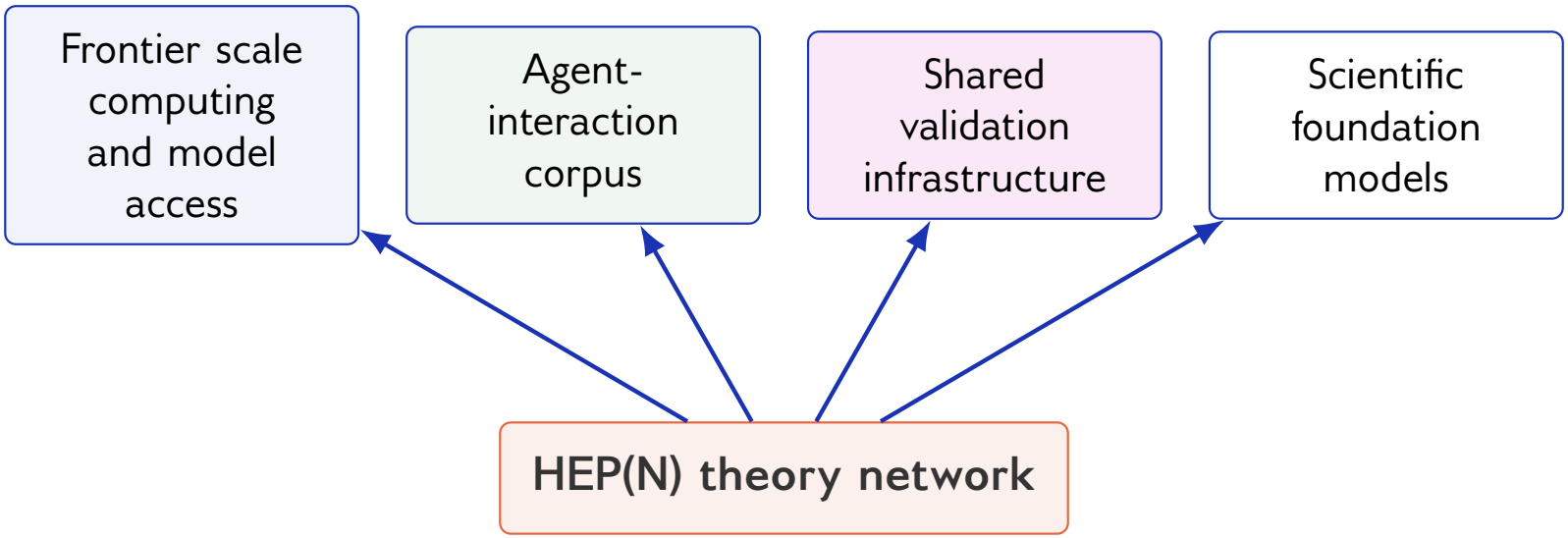
Unless theory does too, we do not get
the frontier resources we need.

Why organize? New technologies mean new capabilities

The methods of particle theory can change, but the practice does not

- What we do can be the same:
  - build models, compute consequences, compare to data
  - study data, infer models, calculate new consequences
- How we do can it change:
  - search literature, make derivations, write code, run calculations, summarize results, (insert your least favorite but necessary scientific task): an agent can do part of each
- What we do can also change:
  - Theory of AI
  - Discover new representations of the natural world
  - *etc.*

Potential assets only a community can have



These assets did not exist five years ago.
Out of reach for one PI,
one group, most single institutions.

Frontier scale computing and model access

- A fair share decided by the community, rather than by which group happened to write which grant.

Agent-interaction corpus

- Long sessions, wrong turns, corrections, and the reasoning that got there: training and evaluation material for scientific agents.
- Pooled across many groups, it can improve our models and interactions with them.

Shared validation infrastructure

- A private benchmark cannot settle claims.
- Standardization: two groups' claims can be put side-by-side.
- Variety: physics and not one group's taste in problems.

Scientific foundation models

- Experiment data is half the picture. Theory data, simulated and derived, is the other.
- Value scales with the diversity of data.

Does Theory have a special place?

Some scientific tasks are just common with experiment

- Optimization, efficiency, and workflow problems likely share common solutions.
- Shared benchmarks and common formats, so results can be compared across the whole field.
- Foundation models built on experiment **and** theory data. Value scales with the diversity of the data.

A model of experiment and theory data together is something neither wing builds alone.

Why theory has a special place

- Theorists generate simulation data, for experiment and for themselves.
- And other sorts of data:
  - equations (trial truth statements)
  - functions (constrained mathematics)
- Theory behind NN itself: connections to QFT and UQ.
- Reasoning trail. Logic: training data.
- Big problems that need to be studied, to guide current research and to decide on new experiments.

Phenomenology techniques and tools used HEP-wide have come from the theory/pheno community.

Theorists work differently

An allocation model built around experiment will not always fit us

- Rubber ducking.
- Long, interactive sessions.
- “Failures” can be good data.

Shared infrastructure with different modes of access.

Why organize? Survive & Thrive

1. Deeper dive into data. Questions we currently do not ask because they cost too many person-hours.

2. Monte Carlo of ideas. Generate and test model space the way we already generate and test event space.

3. Drawing in the best and brightest talent. The students who want to work this way will go where they can.

Theory Community testimonial

First, I am on an experiment (CMS)

- That has helped me discover problems and find solutions, that I would have missed. I have learned and I have taught.

- **Claim:** event generator groups are theory's close equivalent to an experimental collaboration.

Event generator developers model

Shared assets

- git repos
  - work integration
  - builds and validation
  - issue tracking
- purpose and vision
- sustainability plan
- policies
  - AI & coding [link]

With

- local research interests
- projects that can be attached to students and grants

It works!

USQCD pooled HPC and software resources
for the lattice theory community
more than 20 years ago.

SAINT

Self-organizing

AI Network

for Theorists

(Self-organizing, shared, or any good word beginning with S.)

Where to start

1. National Lab theory groups serve as an anchor: organize a meeting ASAP focused on the four assets.

2. A pilot artifact repo on the GMP, open to all registering theorists. Artifacts == human-agent interactions, skills, tools. *

3. Connect to the broader national effort through the GMP, using common formats.

Education, training & workforce development.
Important to get this right.[a]

[a] Let's stop comparing models to "undergrad", "first year graduate student", etc.

*This one first: it is the cheapest to build, and its value grows fastest with the number of groups contributing.

The bottom line

Theorists can keep what is valuable and unique in how they work and still share resources

- Organizing is how we survive the shift in funding, and how we thrive on what the tools now make possible.
- Event generator groups and USQCD show it already works.
- Theory needs a voice in a field-wide AI effort, and that requires a community.

Theory is just forming.
The best time to connect the two wings is now.

Blueprint White Paper

Open for endorsements

Blueprint for an AI-Enabled Particle Theory Community

George T. Fleming[1], Daniel C. Hackett[1], James Halverson[2], T. J. Hobbs[3], Stefan Hoeche[1], Brandon Kriesten[3], Konstantin T. Matchev[4], Tony Menzo[1,4], Stephen Mrenna[1], Benjamin Nachman[5], Prasanth Shyamsundar[1], David Shih[6], Jesse Thaler[7], Jure Zupan[8]

[1] Fermi National Accelerator Laboratory, [2] Northeastern University, [3] Argonne National Laboratory, [4] University of Alabama, [5] SLAC National Laboratory, [6] Rutgers University, [7] Massachusetts Institute of Technology, [8] University of Cincinnati

September 14, 2026

Abstract

Rapid advances in the capabilities and reach of artificial intelligence (AI) are changing research in particle physics. In this new era, AI is an innovative tool for theoretical physicists, while the interaction of AI and particle theory is itself a developing field. We argue that particle theory requires a community-scale effort to realize its full potential. We show how particle theory occupies a unique position in the AI for physics landscape, and describe the complementarity of AI and theory developments. We conclude with a series of recommendations on how to build an AI-enabled theory community for the Genesis Mission.

http://smrenna.github.io/share/whitepaper.pdf

https://indico.fnal.gov/event/75639/surveys/