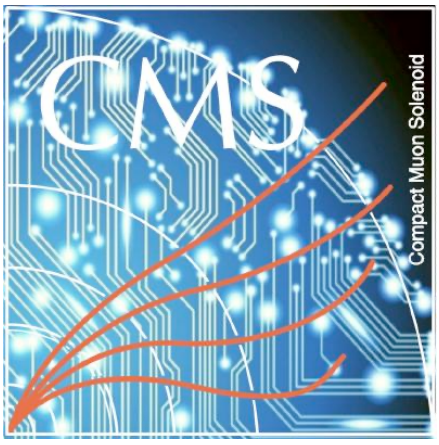
# CMS ML KNOWLEDGE GROUP NEWSLETTER

*August 2026, v4*

Welcome to the fourth edition of the CMS ML Knowledge Group Newsletter!

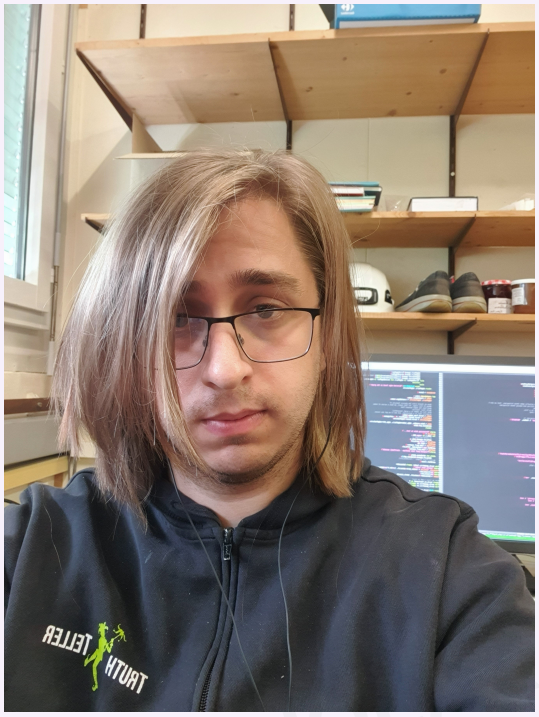
**QUICK PEEK**

- Scholar Spotlight: Andrew Loeliger from Princeton
- Agents in HEP
- Events on our Radar



# SCHOLAR SPOTLIGHT

**Meet Andrew Loeliger, a Postdoctoral researcher at Princeton working on CICADA**




**Q: Could you tell us a little more about your background and how you got involved in the CICADA project?**
**A:** I did my undergraduate degree at the University of Colorado Boulder, where I got my first taste of particle physics as an undergraduate researcher working on simulation for the experiment that is now called DUNE at fermilab (at the time it was still called LBNE). I did my Ph.D with the CMS group at the university of Wisconsin Madison where I was working on the Higgs to taus analysis for Run 2 Data, which was some of the first STXS measurements and the first differential measurements in the decay channel of the higgs. At the end of my Ph.D. I was working on doing boosted Tau ID which is how I first started doing ML, and took to it so much I was learning about it as a hobby. I knew my current PI at princeton via her work on the HIggs to taus analysis, so when I told her I was graduating, she offered me a spot working on CICADA.

**Q: For readers who might not know, what is CICADA and what problem does it aim to solve at the LHC?**
**A:** The simple version is this: analyses at the LHC take one of two approaches to trying to find something novel. In the first approach, you have some standard model process you know is there, and then you attempt to measure it with more precision to attempt to find some deviation between theory and measurement to try and find something new. For the second, you typically have some beyond standard model scenario you want to find, and you set about designing a way to make some signal enriched measurement to find it, if it is at all there. The problem with this second technique is it's biased. You come into it with ideas of what you want to find, and as a result, you are not looking at all the things that could be there. People who work in anomaly detection often bring up the "lost keys/street lamp" analogy which goes something like this: "If you're out late at night and lose your keys, where do you look? You look under street lights or anywhere where there is light, because that's where you can look. But it has nothing whatsoever to do with where your keys might actually be. You're so caught up in the way you know how to search well, you don't focus on expanding the search". This is one thing CICADA (and anomaly detection at the LHC in general) gives us. A wide, less biased search of where interesting things might be. Targeted searches in the manner I have just described won't stop being relevant, but anomaly detection can help to shape where and how we spend our effort on these high impact kind of efforts. Instead of searching everywhere, we might be able to find anomalous regions or places, and help filter down the list of good candidates to take a look at, and help focus the hunt for new physics.

**Q: What is your specific role or contribution within the CICADA collaboration?**
**A:** That's tough to put exact words to, I have

been a little bit of everything when it is called for, but this is true of just about everyone on the project. I've been one of CICADA's public faces for a while (I used to refer to myself as "the ambassador to CICADA"). I developed its emulator software, and other offline software parts, including the bits that unpack and read it's inputs into CMS's offline software. I've done upgrade work for new models, testing ideas for pileup resilience (credit should go to Lino Gerlach here though as the person who came up with the technique we use now). I've done analysis of the outputs of the trigger, trying to characterize what we get, and what kind of characteristics the trigger has in running. I've done a lot of auxiliary machine learning work developing models for the trigger to try and understand how we put thresholds on it to get specific rates at the trigger, and how those change with pileup. Right now, I am trying to organize our effort with a few of the other postdocs who work on the anomaly triggers to publish a paper on this method, writing paper text and making some of our final plots for publication.

**Q: While working on developing CICADA, what has been the biggest challenge so far, and how have you approached solving it?**

**A:** That's tough too, because there hasn't been a single monolithic challenge that sticks out from the rest, it's just been a lot of challenges of similar magnitude. Picking one at random that is at the top of mind right now, I would say commissioning was very difficult. The challenge was not necessarily machine learning related, but a broader hardware, software and machine learning debuging period during commissioning when we were getting ready to turn the whole thing on. CICADA takes 252 inputs, in a very specific geometric pattern, that has to be streamed to the CICADA cards via 36(?) optical links at precise trigger timing constraints. Well, when we first turned it on, the output from the emulator and the hardware didn't agree. This set off a huge series of tests. We had already tested that given a simple pattern we could make hardware and software agree. This was done jointly with the engineers behind a lot of the CICADA hardware and firmware, and we had eventually convinced ourselves that given the same inputs, hardware and emulator in our test setup would agree (as an aside, this is a thing everyone should take away, always test whatever you make, I spend a lot of time telling people at the wonder of unit tests). I had developed some patterns that when put across each individual link gave a different CICADA score each time (I called these "link thumbprints") and eventually we got all these to line up from hardware testing and the emulator. The trick then became that clearly the hardware was not getting the inputs we expected in production. The initial suspect was that the links were not being read into inputs in the right order at the CICADA side. Unfortunately the CICADA hardware didn't save exactly what the inputs it received were and what it thought the order was, but the hardware before it did. So, I sat down to try and figure out if there was some way to reorder the link information I had been given, to recreate CICADA's buggy output and see if we could reverse engineer what the switch was. With 36 links, you have 36! possible combinations of ways they could be sorted, so brute forcing it wasn't an option. I saw down and did some research on "derivative free optimization" (another great tool to be aware of, I wrote both genetic algorithm link swapping tests, and simulated annealing link swapping tests under the assumption that the more links I got right, the closer it would be to CICADA's output), to see if I could come up with a quicker way to swap links around to come up with the right answer. This didn't actually work, but the fact that there wasn't a simple way to swap the links meant that that likely wasn't the problem. Eventually we got into doing some bit level unpacking of the exact outputs, and the engineers had the idea to embed some CICADA metadata into unused bits. I ended up with a lot of information we took that looked like the attached photo. and eventually the engineers figured out that the links were not aligning appropriately and did some firmware revisions. I could go on. Getting the right thresholds on anomaly score to get the right rate was a high pressure

challenge. We had very limited time to try and reoptimize where we set thresholds to not go over trigger rate budgets. I ended up using a mix of three different ML methods to map scores to rates (plus the statistical practice of bootstrapping, yet another great tool to be aware of), then map PU conditions to a multiplicative rate factor from our basic score->rate model, and then another to correct from that model to the few actual rate points we had online at a given pileup. Traditional ML models like gaussian processes are still great tools for stuff like this, it doesn't all have to be neural nets.

**Q: CICADA heavily relies on machine learning and fast decision-making for real-time triggers. What are some of the innovations your team has implemented to make this possible?**
**A:** Good question. There are two things that make this possible. The first is "knowledge distillation". Knowledge distillation is a well explored ML technique of training a smaller more model on labels provided by a larger one, to try and compress the relevant characteristics into a lighter overall model In our case it also cut out an extra step of the anomaly detection process (going from reconstructed image to error). The second is High Level Synthesis, specifically a package called HLS4ML, which synthesizes neural net models into c++ code that can then be made into FPGA firmware. We are not the authors of this package, but one of the first major users, but I would be remiss to not give the HLS4ML authors a huge amount of the credit for making this idea work.

**Q: For researchers and students who want to get involved in projects like CICADA, what advice would you give them?**
**A:** I guess that depends a bit on how serious you want to be. If you want to explore and learn, I guess my advice is just to experiment with stuff, it doesn't have to be practical. For myself, I started off just reading up on the fundamentals, and then just continually trying out new models, new ideas, new whatever with whatever bits of data I had hanging around. That's one of the great bits of being on the LHC is you don't lack for data at various levels of processing. And you learn a lot even on impractical projects. Even right now I am testing ideas for different ways we could make CICADA work better, and be more noise and conditions resilient, and I've been exposed to whole new ideas even in the relatively small world of just auto-encoders, unsupervised learning and anomaly detection. If you want to get involved in a very serious project very quickly my recommendation is planning, especially with regards to what problem you are going to solve, and how you are going to assess how well it is being solved. It is very easy in the machine learning world to come up with nebulous problem statements, and you will get a lot of false starts, and trivial, uninformative modeling this way. You'll also get problems that are not just one model problems, sometimes it requires a whole system of models to be thought out. CICADA was thankfully a very clear concept from the start, but even then we have had to do some thinking about how we quantify how well CICADA is doing. A thing you should catch in planning, is to try and find where simple or trivial solutions work better. It is good practice to start trying to solve every artificial intelligence/machine learning problem first with non-artificial intelligence, and it gives you a benchmark to test against. If and when you come up against problems where that is clearly just not good enough, that's how you get your candidate for a good ML project.

**Q: What excites you most about the potential impact of CICADA on future LHC runs or particle-physics searches?**
**A:** One of the things that excites me about CICADA and it's impact, is that it means that this kind of modelling is present at every layer of the detector from triggering and data acquisitions all the way through to reconstruction and analysis level. "By hand" solutions can't and shouldn't disappear for reasons that I have discussed, but I hope I have said the word "model" enough to impress upon everyone that I believe in the power of clever modelling to really increase what the optimum is in almost every way the detector operates. The other good thing that I am excited about is that this is unsupervised

machine learning. For a while there it was easy to classify ML as just a way we make particle ID taggers, but this is a whole new paradigm, this is a new way to do ML for experiments, and it is minimally biased. In my opinion minimally biased/model independent (yes I appreciate the irony of the term ”model independent” here) searches and techniques don’t get the kind of attention that they really should, because these should be out in front leading the way for more targeted/model dependent ones.

**Q: Outside of research, what do you enjoy doing in your free time?**
**A:** I enjoy art, specifically pencil or pen and ink sketching, but I really regret that I don’t make as much as I used to. These days I spend a lot of time messing with ML ideas outside of work as well, just trying stuff. Doing ML research for physics there are very few places where you would run into reinforcement learning (or perhaps I just hear about it less) but it’s a really cool area of machine learning and ”find an optimal policy” is a pretty useful thing to be able to do generally, so I have been spending time experimenting with that. Most ML techniques are designed to give you a single answer, a central value, but most things in life are probable distributions or at the very least come with uncertainties, and there are whole fields of machine learning out there that try to implement probability distributions as outputs (outputting mu and sigma for a normal distribution for example) or try to implement bayesian ideas into machine learning (I for one am shocked that Bayesian Neural Networks are not a bigger thing). There are heaps of bayesian modelling ideas even outside of neural networks I am surprised it has taken me so long to learn about and find uses for, Gaussian Processes or Bayesian hierarchical modelling seems like things with a pretty stunning amount of use cases. I don’t want to sound silly, but I think what experimenting with ML has shown me, is that I am fundamentally a tool and model maker, and I am happiest in physics when I get to be making tools and models.

*Author: Ellison Scheuller*

**Summary from Bunches of Agents: Mini talk series (part 1)**

CMS's ML Innovation team started a new talk series on June 22nd called "Bunches of Agents," looking at how AI agents might fit into particle physics research.

Gordon Watts (University of Washington) tested GPT-5.4-mini and GPT-5.5 live during his talk, asking them to write a full ATLAS top-mass analysis from scratch. The result was usable after he checked it by hand, but the model made its own choices about object-selection cuts without saying so, so the same prompt didn't always produce the same plot twice. Watts talked through the main failure modes he's run into: hallucination, lack of grounding in real experiment data, and non-determinism. His main suggestion was to give agents direct access to real dataset catalogs and trigger databases, rather than relying on the model to guess that information.

Philip Harris (MIT) and collaborators, frustrated by how long a normal LHC analysis takes to clear review (sometimes 2-3 years), decided one day to just let a team of coordinating AI agents loose on one. The framework they built, called JFC ("Just Furnish Context"), is described in their paper "AI Agents Can Already Autonomously Perform Experimental High Energy Physics." Twenty-four hours later they had a working refit of the ALEPH Z lineshape. From there they ran JFC on ten full reanalyses of old LEP and CMS measurements, including the Lund jet plane density, a measurement nobody had actually done back in the LEP era. Reanalyzing CMS's published Higgs-to-tau-tau result, JFC landed on a signal strength of 1.20 ± 1.13, versus the real published 1.01 ± 0.59, in the ballpark but noisier. Harris was upfront that this isn't publication-quality, and that the reaction online was split down the middle: real excitement next to Slack messages calling it "so bad" and a viral post calling it terrifying. His bigger point was about pace: on a $300-a-month budget, that same team could now redo all ten analyses in about a week, and the newer models coming out are closing the gap fast.

A related workshop at ETH Zürich three days later covered similar ground with a wider group of speakers: Daniel Murnane, Fazl Barez, Eric Moreno, and Sofia Palacios Schweitzer, under the title "English is the new coding language." Schweitzer presented COLLIDER-BENCH, a new benchmark for measuring how well agents can reproduce a physics analysis.

Across both events, the discussion of AI agents in HEP has moved from whether they can contribute to analysis work at all toward how their output should be validated and integrated into existing research workflows.

*Author: Ellison Scheuller and Claude ;)*

## EVENTS ON OUR RADAR

→ Fast Machine Learning for Science Conference 2026

## JOIN US

This newsletter was brought to you by the **CMS ML Knowledge Sub-group**. We meet every two weeks and **welcome new members**!

If you've enjoyed this newsletter, please let us know. If you have an idea for content, want to nominate someone for an interview, or contribute, we'd love to hear from you.


**EDITOR**

Ellison Scheuller

**CONTACT**

Daniel Diaz & Marius Snella Köppel

cms-conveners-ml-knowledge@cern.ch