Oz Amram

oz.amram@gmail.com • github.com/OzAmram • Google Scholar • ozamram.github.io

Machine learning + particle physics researcher at Fermilab. 18 publications, 701 total citations. Led one of the first AI-based anomaly-detection searches in particle physics. Wrote CaloDiffusion, a denoising diffusion model for detector simulation, from scratch. Multiple other projects on anomaly detection, applications of generative models, and foundation models for particle physics, each with public codebases.

Research Positions

Wilson Fellow & Associate Scientist, Fermilab 2026–present

- Research on machine learning for particle physics; member of the CMS experiment at the Large Hadron Collider (CERN).
- Co-convener, CMS exotic-searches-with-jets analysis group (EXO Jets+X), 2024–present: review and manage $\sim$15 active analyses (papers), 16 public results during tenure.

Postdoctoral Researcher, Fermilab 2022–2026

- **Lead analyzer**, first CMS paper using ML-based anomaly detection to search for new particles: five complementary unsupervised methods applied to 30M collision events ($\sim$5 TB), with up to sixfold gains in discovery sensitivity over traditional methods across a wide range of signal types never seen in training.
- **Designed the search's statistical and interpretability frameworks** — the first for an anomaly-detection search at the LHC — to characterize what makes a flagged event anomalous and map it back to physical detector signatures.
- Wrote **CaloDiffusion** from scratch (architecture, noise schedule, sampler): denoising diffusion for calorimeter simulation, best-in-class fidelity in the community CaloChallenge benchmark and up to $1000\times$ faster than the Geant4 physics simulation. Trained on 500k showers in $\sim$100 GPU-hours.
- Leading the first application of generative models to simulation in CMS: a generative model with 500,000-dimensional output for the High-Granularity Calorimeter, trained on 1.5M example showers in $\sim$300 GPU-hours. Built the evaluation framework comparing multiple models against quantitative quality metrics.
- Developed a method for calibrating the simulation of high-dimensional sprays of particles (jets) and assigning uncertainties to it, now a standard technique in CMS and used by 5+ analyses.
- Mentored graduate students at UCSD, UCSB, Cornell, KIT and Hamburg on anomaly-detection projects.

Education

Ph.D., M.A. Physics, Johns Hopkins University 2016–2022

- Thesis: *Searching for Anomalies in Proton-Proton Collisions at the Large Hadron Collider*. Advisors: Morris Swartz, Petar Maksimovic.
- Independently developed Tag N' Train, an anomaly-detection algorithm and a top performer in the LHC Olympics community competition.

B.S. Physics, minor Philosophy, Carnegie Mellon University 2012–2016

Selected Publications

h-index 9, with 11 of the 18 publications as primary author or lead contributor; 423 citations and h-index 8 excluding self-citations. Additionally an author on 400+ CMS collaboration papers.

1. O. **Amram**, M. Letizia, and M. Kuusela. "Model-Agnostic Signal Discovery with Machine Learning: Bridging the Gap Between Theory and Practice" (2026). arXiv: 2605.31103

   Lead author. Review paper on anomaly detection, covering the statistical basis, applications and validation strategies. Invited contribution to a series on AI in particle physics.

2. CMS Collaboration. "Machine-learning techniques for model-independent searches in dijet final states". *Sub. to Mach. Learn. Sci. Tech.* (2025). arXiv: 2512.20395

   Second lead author. Contributed the weakly supervised anomaly-detection results across all performance studies, and ran the rediscovery validation of boosted top quarks — recovering a known rare process without labels in CMS data as an end-to-end test of the method.

3. CMS Collaboration. "Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s}$ = 13 TeV". *Rept. Prog. Phys.* (2025). arXiv: 2412.03747

   Lead author. Five complementary anomaly-detection methods, with up to sixfold gains in discovery sensitivity over traditional methods. Designed the statistical and interpretability frameworks, the first for an anomaly search at the LHC. Selected as an IOP Research Highlight.
4. O. **Amram** and M. Szewc. "Data-Driven High-Dimensional Statistical Inference with Generative Models". *JHEP* (2025). arXiv: 2506.06438

   Lead author. Replaces binned histogram fits with a learned generative model, enabling likelihood fits in 5+ dimensions; $\sim$40% improvement on a benchmark measurement. Code: github.com/OzAmram/HI-SIGMA.
5. O. **Amram** et al. "Aspen Open Jets: unlocking LHC data for foundation models in particle physics". *Mach. Learn. Sci. Tech.* (2025). arXiv: 2412.10504

   Built and released the dataset and its processing tooling: 180M jets (200 GB) derived from public collider data, to support pretraining of foundation models on real rather than simulated data. Code: github.com/OzAmram/AOJProcessing.
6. CMS Collaboration. "A method for correcting the substructure of multiprong jets using the Lund jet plane". *JHEP* (2025). arXiv: 2507.07775

   Analysis contact. Method for correcting and assigning uncertainties to simulated jet models where simulation and data disagree — a calibration problem for classifiers trained on simulation. Now a standard technique in CMS. Code: github.com/OzAmram/LundReweighting.
7. O. **Amram** and K. Pedro. "Denoising diffusion models with geometry adaptation for high fidelity calorimeter simulation". *Phys. Rev. D* (2023). arXiv: 2308.03876

   Lead author; wrote the architecture, noise schedule and sampler from scratch. Geometry-adapted denoising diffusion over high-dimensional structured data; achieved best-in-class sample fidelity under the physics-based evaluation metrics of the community CaloChallenge benchmark. Code: github.com/OzAmram/CaloDiffusion.
8. O. **Amram** and C. M. Suarez. "Tag N' Train: a technique to train improved classifiers on unlabeled data". *JHEP* (2021). arXiv: 2002.12376

   Independently developed. Co-training scheme for particle physics data that learns from noisy, mixed-sample supervision with no labeled examples of the target class. Code: github.com/OzAmram/TagNTrain.

Technical Skills

- **Languages:** Python (primary, expert). C++ (used heavily through 2022, not recently).
- **Frameworks:** PyTorch (primary); TensorFlow/Keras.
- **Methods:** diffusion models; weakly supervised, co-training, autoencoders and unsupervised methods for anomaly detection; generative models for fast simulations and likelihood-based inference; uncertainty quantification.
- **Released code:** CaloDiffusion, CMS HGCaloChallenge, HI-SIGMA, LundReweighting, AOJProcessing, TagNTrain (github.com/OzAmram).

Awards & Funding

- **DOE Genesis Mission grant** (2026): *AIDA-Scout*, real-time anomaly detection on the full LHC collision rate. Originated the idea and wrote majority of the successful proposal; senior/key personnel. $135k Fermilab share of $660k across five institutions, nine months from August 2026.
- Fermilab LHC Physics Center Distinguished Researcher (2026).
- Rising Star in Experimental Particle Physics, University of Chicago (2021).

Additional

- **Talks:** 21 conference presentations (7 invited); 11 invited seminars and colloquia, including MIT, UChicago, Stanford Linear Accelerator (SLAC), Michigan, Purdue, UIUC and TRIUMF.
- **Teaching:** Founding organizer of the annual Machine Learning for Fundamental Physics school ($\sim$40 in person, $\sim$100 virtual); gave the anomaly-detection lecture and ran the tutorial.
- **Writing:** $\sim$20 articles for Particle Bites explaining recent particle physics results to a general audience; wrote the CMS and CERN articles on the anomaly-detection search.
- **Reviewing:** NeurIPS Machine Learning for Physical Sciences workshop (2025–6); ACAT (2023). Journal referee for SciPost Physics, Physical Review D, Physics Letters B, Scientific Reports.