RODEM Scientific Report

François Fleuret, Slava Voloshynovskiy, Tobias Golling, André Csillaghy

December 6, 2025

1 Project Metadata

- Grant No. : 193716
- Title : Robust Deep Density Models for High-Energy Particle Physics and Solar Flare Analysis (RODEM)
- Funding scheme : Synergia
- Institute : Département d'Informatique, Université de Genève
- PIs :
  1. François Fleuret, Centre Universitaire d'Informatique Université de Genève, Switzerland
  2. André Csillaghy, Fachhochschule Nordwestschweiz Hochschule für Informatik, Switzerland
  3. Tobias Golling, Département de physique nucléaire et corpusculaire Université de Genève, Switzerland
  4. Slava Voloshynovskiy, Stochastic information processing, Campus Battelle Université de Genève, Switzerland
- Period : 01.12.2020 – 30.04.2025

2 Executive Summary

Objectives. The RODEM project aimed to develop large-scale machine learning methods with provable guarantees for rare event detection and prediction in High Energy Physics (HEP) and solar astronomy. Our core objectives were to create: (1) improved forecasting tools trainable from large, high-dimensional datasets with limited supervision, (2) computationally efficient generative models as surrogates for expensive simulators, and (3) anomaly detectors with formal guarantees for super-rare event regimes.

Major Results. We solved the critical "mass sculpting" problem in particle physics anomaly detection through a systematic progression from CURTAINs morphing methods to the New Physics Learning Machine (NPLM), achieving deployment within the ATLAS collaboration for active new physics searches. Our generative modeling advances created a complete pipeline from individual particle clouds to full event generation (PIPPIN), representing the first transformer-coupled diffusion model for variable-multiplicity events—crucial for High-Luminosity LHC simulation. Novel theoretical contributions include neural samplers based on continuity equations and energy-based diffusion models for free-energy estimation, achieving two-orders-of-magnitude speedup for point cloud computations. Computational efficiency breakthroughs in sparse attention, Mamba-based architectures, and DeepEMD directly address large-scale physics analysis bottlenecks. Cross-domain validation through SkyCURTAINs successfully adapted collider methods to Gaia astrometry, demonstrating broader applicability beyond our original scope. In solar astronomy, we increased the reliability of the modeling of solar flare prediction, laying the grounds to a new foundational model based on the data from the Solar Dynamics Observatory (SDO).

Significance and Impact. This work directly addresses two central challenges limiting discovery potential at the LHC: model-dependent search bias and prohibitive simulation costs. Our anomaly detection methods enable model-agnostic discovery searches, while our generative models provide scalable simulation alternatives. The progression from proof-of-concept to analysis-ready tools with demonstrated ATLAS deployment represents a paradigm shift toward data-driven physics discovery. Complementarily, in solar astronomy, the modeling of solar activity is a challenge that cannot be completely understood analytically, partly due to the chaotic nature of some of the involved phenomena. Therefore, the data driven approach adequately provides modeling material through the data processing, taking advantage of the very large amount of observations available. Our work has shown that we are able to increase reliability of the existing model by providing suitable foundations for downstream applications. The collaboration resulted in 25+ publications including top-tier venues, with methods adopted by major experimental collaborations.

3 Work performed (methods & progress)

The discovery potential of the Large Hadron Collider (LHC) is increasingly limited by two challenges: the risk that model-dependent searches miss unforeseen signatures, and the prohibitive computational cost of high-fidelity Monte Carlo simulation. This program addresses both through advances in **weakly supervised anomaly detection** and **deep generative modelling**, with demonstrated transfer to astrophysical data. Together, these contributions move from proof-of-concept methods to statistically principled, analysis-ready tools for high-energy physics. In the field of solar astronomy, similar approaches, methods and models can be applied to increase the (numerical) understanding of solar activity, especially in the context of active region dynamics and solar flares.

3.1 Anomaly Detection for New Physics Discovery

In **Anomaly Detection**, the work began with variational autoencoder (VAE) based anomaly detection approaches. While effective in capturing high-dimensional features, this line of research revealed a critical limitation: strong correlations between learned anomaly scores and the jet mass variable led to "mass sculpting," thereby reducing discovery sensitivity. This issue was systematically analyzed in (Golling et al., 2023c) and presented at NeurIPS 2020. R

ecognizing these limitations motivated a shift toward template-based approaches that explicitly model and control mass correlations. The first such framework, *CURTAINs for your sliding window* (Raine et al., 2023), employed optimal transport to morph background sidebands into the signal region, enabling data-driven background templates. In 2024, *CURTAINs Flows for Flows* (Sengupta et al., 2024a) advanced this approach using conditional normalizing flows for likelihood-based morphing, while *DRAPES* (Sengupta et al., 2024c) extended the method to significantly more observables by employing diffusion models. A simulation driven approach was also developed using a similar framework (Golling et al., 2023a) based on a generic framework that the group developed for calibration (Golling et al., 2023b). The CURTAINs approach was also deployed within the ATLAS collaboration to search for new physics (Collaboration, 2025). Improvements, variations and speed-ups of the methods were proposed in (Leigh et al., 2024a; Sengupta et al., 2024b; Oleksiyuk et al., 2024a). Methods were compared in (Golling et al., 2024b).

Most recently, the *New Physics Learning Machine (NPLM)* (Grosso et al., 2025) formalized resonant anomaly detection as density-ratio estimation, yielding improved stability and sensitivity at very low signal fractions. Together, these contributions chart a clear progression from VAE-based unsupervised methods to statistically rigorous, template-driven pipelines with demonstrably higher sensitivity.

3.2 Generative Models and Fast Simulation

In **Generative Modelling**, the 2023 publications laid the groundwork. *PC-JeDi* (Leigh et al., 2024b) introduced diffusion models for particle-cloud jets, while *EPiC-ly Fast* (Buhmann et al., 2023) proposed both faster diffusion (EPiC-JeDi) and the first permutation-equivariant flow-matching generator (EPiC-FM). In 2024, *PC-Droid* (Leigh et al., 2024c) accelerated diffusion-based jet generation using improved ODE solvers and distillation, achieving practical speeds with high fidelity. Extending beyond jets, the 2025 *PIPPIN* (Quétant et al., 2024) model generated full events with variable multiplicities using transformers coupled to diffusion and flow-matching, an important milestone toward end-to-end ML-based simulation. Generative models are also used for variational inference for pile-up removal at hadron colliders with diffusion models (Algren et al., 2025b). Other HEP applications of generative models and optimal transport include powerful object reconstruction techniques (Leigh et al., 2023; Raine et al., 2024), novel calibration methods (Algren et al., 2023, 2025a) and their application in ATLAS (Aad et al., 2025),decorrelation (Klein and Golling, 2022; Algren et al., 2024) and generalization (Rothen et al., 2025).

Calorimeter Simulation HEP analyses rely on comparisons (Aad et al., 2010) of real data generated through experiments with reference data generated through simulations driven by physics theory and the underlying test hypotheses. Detector simulation is very compute intensive and the compute needs for future detectors is expected to exceed the available budget. Calorimeter simulation is one of the most expensive constituents of detector simulation. Deep learning methods have gained popularity in high energy physics for fast modeling of particle showers in detectors. Detailed simulation frameworks such as the gold standard GEANT4 (Agostinelli et al., 2003) are computationally intensive, and current deep generative architectures work on discretized, lower resolution versions of the detailed simulation. The development of models that work at higher spatial resolutions is currently hindered by the complexity of the full simulation data, and by the lack of simpler, more interpretable benchmarks.

In our work (Sinha et al., 2023) we propose SUPA, the SUrrogate PArticle propagation simulator, an algorithm and software package for generating data by simulating simplified particle propagation, scattering and shower development in matter. SUPA achieves extremely fast generation and is easy to use compared to GEANT4, but still exhibits the key characteristics and challenges of the detailed simulation. SUPA generates thousands of particle showers per second on a desktop machine, a speed up of up to 6 orders of magnitudes over GEANT4, and stores detailed geometric information about the shower propagation. SUPA provides much greater flexibility for setting initial conditions and defining multiple benchmarks for the development of models. Moreover, interpreting particle showers as point clouds creates a connection to geometric machine learning and provides challenging and fundamentally new datasets for the field.

Flowification The two key characteristics of a normalizing flow (NF) is that it is invertible (in particular, dimension preserving) and that it monitors the amount by which it changes the likelihood of data points as samples are propagated along the network. Recently, multiple generalizations of NFs have been introduced (Nielsen et al., 2020; Huang et al., 2020) that relax these 2 conditions. On the other hand, neural networks (NNs) only perform a forward pass on the datapoint, there is neither a notion of an inverse of a neural network nor is there one of its likelihood contribution.

In our work (Máté et al., 2022) we argue that certain NN architectures can be enriched with a stochastic inverse pass and that their likelihood contribution can be monitored in a way that they fall under the generalized notion of NF mentioned above. We term this enrichment *flowification*. We prove that neural networks only containing linear and convolutional layers and invertible activations such as LeakyReLU can be flowified and evaluate them in the generative setting on image datasets.

3.3 Theoretical Foundations and Sampling Methods

Neural Samplers Data-free sampling from probability distributions specified by an energy function is a long standing problem in the natural sciences. While traditional methods such as Markov Chain Monte Carlo (MCMC) or Molecular Dynamics (MD) provide theoretical guarantees in the limit of infinite samples, their long autocorrelation time poses considerable challenges when sampling multimodel target densities with large energy barriers between the modes.

In Máté and Fleuret (2023) we propose to sample by learning a one-parameter family of interpolating distributions between the base and the target distibutions and enforce the continuity equation along this interpolation. We demonstrate our objective is less prone to mode collapse then other the standard, reverse-KL based training of neural samplers.

In Mate and Fleuret (2024) we connect the multi-scale nature of lattice field theories with that of neural operators. We demonstrate that with a careful design of the architecture, initially learning to sample at coarser resolutions and incrementally working on finer lattices provides considerable computational speedup over working directly on a larger lattice.

Free-energy Estimation with generative models Estimating free-energy differences is at the heart of understanding a wide range of physical, chemical, and biological processes, from protein folding and ligand binding to phase transitions in materials. These calculations offer invaluable insights into the stability of molecular conformations, the spontaneity of chemical reactions, and the mechanisms that drive phase changes.

In Mate et al. (2024) we show that an energy-based parametrization of diffusion models allows for free-energy estimation via thermodynamic integration (TI). We demonstrate the approach on sampling the grand canonical ensemble and accurately estimating the excess chemical potential of a Lennard-Jones fluid. In Máté et al. (2025) we extend this method to the estimation of solvation free energies and predict the hydration free energies of rigid water and methane in good agreement with experimental measurements.

Deformations of Boltzmann Distributions Sampling from unnormalized densities has been studied by many due to its relevance for the sciences (Noé et al., 2018; Boyda et al., 2021). The problem can be summarized as follows. Given an energy function (unnormalized negative log-density) $f : \mathbb{R}^n \to \mathbb{R}$ can we efficiently generate samples from the probability density $p(x) = \frac{1}{Z}e^{-f(x)}$? In particular, there are no samples given, all we have is the ability to evaluate f for any sample candidate. A popular strategy to attack this problem is to use a normalizing flow to parametrize a distribution q_θ and to optimize the parameters θ to minimize the reverse KL-divergence $KL(q_\theta, p)$.

In our work (Máté and Fleuret, 2022) we derive a family of energy functions f_t containing f_0 that defines the ϕ^4 lattice field theory on a two-dimensional lattice. These energy functiosn that define a family of unnormalized densities $p_t(x) = \frac{1}{Z_t}e^{-f_t(x)}$. We prove that these sampling from any member of this family is equivalent to sampling from p_0 in the sense that samples from p_t can easily be transformed to samples from p_0. Moreover, we experimentally find a particular value of τ such that normalizing flows perform better at learning p_τ then at learning p_0.

3.4 Computational Efficiency and Scalability

As our methods matured toward practical deployment, computational efficiency became paramount. This led to several breakthrough contributions that extend well beyond physics applications.

DeepEMD The earth movers distance (EMD), also known as Wasserstein distance is a distance between distributions that is defined as the minimum total of mass-time-distance displacement needed to transform one distribution to the other. In the case of uniform distributions over a finite number of points, it turns into a distance between point clouds that corresponds to finding the one-to one matching that minimizes the sum of the distances between pairs of matched points. Since there is no inherent ordering in point cloud data, computing the EMD between two point clouds involves finding a matching based on the euclidean distance between points. The EMD is the most commonly used distance metric on point clouds, and is extremely useful in many different contexts including high energy physics, medical imaging, astrophysics, computer vision, as well as machine learning applications such as evaluating generative models and comparing probability distributions.

In Sinha and Fleuret (2024) we propose an attention-based model to compute an accurate approximation of the EMD that can be used as a training loss for generative models. To get the necessary accurate estimation of the gradients we train our model to explicitly compute the matching between point clouds instead of EMD itself. We cast this new objective as the estimation of an attention matrix that approximates the ground truth matching matrix. Experiments show that this model provides an accurate estimate of the EMD and its gradient with a wall clock speed-up of more than two orders of magnitude with respect to the exact Hungarian matching algorithm and one order of magnitude with respect to the standard approximate Sinkhorn algorithm, allowing in particular to train a point cloud VAE with the EMD itself. Extensive evaluation show the remarkable behaviour of this model when operating out-of-distribution, a key requirement for a distance surrogate. Finally, the model generalizes very well to point clouds during inference several times larger than during training.

Faster Causal Attention Over Large Sequences Through Sparse Flash Attention In Pagliardini et al. (2023) we introduce fast attention methods and kernels that extend FlashAttention to *dynamic sparse* patterns while preserving exactness and causal masking without materializing the full attention matrix. We support (i) query/key sparsification and (ii) hash/bucketed attention (via LSH-style bucketing) and show strong speedups on long sequences. Our approach uses IO-aware, Triton-based kernels for dynamic sparsity with non-triangular causal submasks, avoiding dense recomputation. Throughput improves with sequence length and sparsity; on long-context language model pretraining we match dense quality at lower wall-clock time. The implementation maintains drop-in compatibility with standard Transformer blocks and training pipelines.

The Mamba in the Llama: Distilling and Accelerating Hybrid Models In Wang et al. (2024) we develop methods to distill strong Transformer LLMs into linear SSM architectures such as Mamba, preserving quality and achieving better inference throughput. Specifically, we distill Transformer LLMs into *hybrid Mamba↔Attention* models (and pure Mamba variants), preserving chat and reasoning quality while enabling faster decoding. Hybrids with reduced attention (e.g., 50–12.5%) achieve near-parity quality with significantly cheaper inference. We further develop a hardware-aware speculative decoding pipeline with fused kernels so hybrids act as efficient verifiers on modern GPUs. Our speculative decoding approach with fused recompute+decode+KV-cache kernels yields consistent speedups across model sizes. The approach works with both instruction-tuned and base LLMs and integrates with common serving stacks.

Understanding and Minimising Outlier Features in Transformer Training In He et al. (2024) we study the emergence of *Outlier Features* (heavy-tailed activations) that harm stability and quantization in Transformer LLMs. We propose the *Outlier-Protected (OP) block*, which removes normalization layers while preserving signal propagation via residual downscaling and attention-entropy regulation (e.g., QK normalization or soft-capping). The OP block formalizes this approach by removing Pre-Norms, scaling residuals $\beta = \mathcal{O}(1/\sqrt{d})$, regulating attention entropy, and optionally scaling MLP inputs. This design substantially reduces activation kurtosis across early and mid layers over long training runs, leading to improved quantization robustness. OP maintains comparable perplexity to strong Pre-Norm baselines, with ablations highlighting the critical role of entropy regulation. Our models, with reduced outliers, perform much better when quantized.

Thinking Slow, Fast: Scaling Inference Compute with Distilled Reasoners In Paliotta et al. (2025) study methods to distill pure and hybrid Mamba models from pretrained Transformers for inference-time compute. Trained on only 8 billion tokens, our distilled models show strong performance and scaling on mathematical reasoning datasets while being much faster at inference for large batches and long sequences. Despite the zero-shot performance hit due to distillation, both pure and hybrid Mamba models can scale their coverage and accuracy performance past their Transformer teacher models under fixed time budgets, opening a new direction for scaling inference compute.

M1: Towards scalable test-time compute with mamba reasoning models In Wang et al. (2025) we introduce *M1*, a Mamba-based reasoning model distilled from a strong Transformer reasoner, then finetuned on math/CoT data and further improved with reinforcement learning. Because M1 generates long chains and large batches efficiently, it enables test-time scaling with markedly higher throughput than same-size Transformers. M1 achieves higher tokens-per-second and lower latency than comparable Transformer baselines on long-chain generation and large batches. Longer RL training sequences lead to higher reasoning accuracy on math and logic benchmarks, demonstrating clear scaling trends. Our practical pipeline follows a distill $\rightarrow$ SFT (reasoning) $\rightarrow$ RL approach with efficient long-sequence batching and serving.

Leveraging the True Depth of LLMs In González et al. (2025) we introduce a *Layer Parallelism (LP)* technique for transformers that executes *pairs* of consecutive decoder layers in parallel under tensor parallelism. Our method restructures attention/FFN sub-blocks so each layer in the pair runs on its own TP shard(s) and the residuals are combined with a single reduce, *halving* inter-GPU synchronizations versus the standard sequential layout. This approach effectively reduces model depth by replacing stretches of sequential layers with contiguous 2-parallel blocks (e.g., reducing depth from 32 to 25 layers) while preserving quality. The method requires no retraining to work and yields consistent end-to-end inference gains across prefill, autoregressive, and 1-token latency scenarios, with improvements growing with longer sequences and larger LP budgets. LP maintains strong perplexity and ICL performance at moderate LP budgets: general ICL tasks retain 95–99% of baseline quality, while math-focused tasks degrade earlier but can be substantially recovered via lightweight fine-tuning of only the parallelized layers. We demonstrate the approach on Llama 2 7B, Llama 3.2 3B, and Qwen3 (4B/14B), indicating robustness across model families and scales.

Fourier Convolutional Decoder: Reconstructing Solar Flare Images via Deep Learning Selcuk-Simsek et al. (2025) introduces, for the first time, an automatic image reconstruction technique for solar flares directly from Fourier-domain signals using deep learning. The model enables large-scale image reconstruction, processing 10,000 solar flare images in about two seconds. The FCD model is implemented as a compact autoencoder occupying only $\sim$11 MB. Its architecture was derived through efficient neural network design principles and subsequently fine-tuned via systematic hyperparameter optimization for the specific task of Fourier-based image reconstruction. Regarding both reconstruction accuracy and computational speed, FCD performs on par with the widely used CLEAN algorithm, while providing superior scalability in terms of throughput and flexibility across diverse imaging setups.

3.5 Cross-Domain Extensions

Beyond High Energy Physics, in **astrophysics**, *SkyCURTAINs* (Sengupta et al., 2025) adapted collider weak-supervision methods to Gaia astrometry, recovering stellar streams such as GD-1 with higher purity than previous approaches. This demonstrates that collider-inspired anomaly detection principles generalize to billion-object astronomical surveys, providing both validation and stress tests for robustness.

Transformers for Graphs Transformers (Vaswani et al., 2017) are extremely powerful architectures that deal with sets and sequences. Transformers have been applied with success to graph-structured data (Dwivedi and Bresson, 2020). Graph data provides a fascinating modality to explore, as graphs can be found anywhere: molecules, social networks, images, and even the World Wide Web. However, due to their computational complexity, Transformers cannot be applied to large graphs, such as social networks or large biochemical compounds. This limitation also makes it hard to capture long-range dependencies between entities in large graphs. In works that are yet to be submitted for publication, we have been exploring two orthogonal axes with the aim of scaling Graph Transformers. One direction of inquiry is the development of efficient Transformer variants that work sensibly on graph-structured data. At the same time, we have been working on the development of retrieval-based systems to help Transformers bypass their computational bottleneck and capture long-range context in the input sequences.

Forward Forward Algorithm for Graphs The Forward Forward algorithm (Hinton, 2022) was introduced at the 2022 Neurips conference by Geoffrey Hinton. It provides a framework for training neural networks without backpropagation, in what is supposed to be a more biologically-inspired regime. Right after the original presentation, we started to explore the issue and work on an extension of this algorithm to graph-structured data, with the aim of submitting the work to conferences in 2023. The extension is non-trivial, as many assumptions that hold for image data, which was tested in the original paper, don't hold anymore on graphs. However, preliminary experiments show promising results in generalizing the frameworks to graph datasets.

3.6 Foundation models for high energy physics

Foundation models are one of the most important modern paradigms of machine learning. One of the first of these models that was developed for high energy physics was the Masked Particle Model (Golling et al., 2024a). This model was shown to significantly improve in dataset size efficiency and training time across many tasks and datasets, paving the way for modern machine learning in high energy physics. Recent work has significantly improved on this (Leigh et al., 2025) further improving the utility of the approach. This work helped kick-start a community effort (Barman et al., 2025; Caron et al., 2025).

3.7 Generative Modelling and Self-Supervised Learning

Building upon the program's efforts in deep generative modelling, our team extended information-theoretic frameworks to unify and improve representation learning and data generation. The TURBO framework introduced in Quetant et al. (2021) and extended in Quétant et al. (2023) provided a principled foundation for analysing and generalising auto-encoding methods, reconciling information bottleneck concepts with multi-representational data. This formulation established TURBO as a reference point for both theoretical understanding and practical applications in astronomy, high-energy physics, and beyond.

Additionally, InfoSCC-GAN (Kinakh et al., 2021a) combined contrastive encoders (InfoNCE), attribute supervision, and EigenGAN-style generators to enable conditional generation with controllable latent factors. The information-theoretic objective ties sufficiency between data–latent and latent–output interfaces. Results on AFHQ and CelebA show improved attribute control and image quality relative to EigenGAN baselines, with targeted ablations on discriminators and losses.

We have also addressed a problem of generation on discontinuous manifolds via continuous stochastic non-invertible networks. (Drozdova et al., 2021) addresses distributions supported on unions of submanifolds/discontinuous supports by learning a contrastive latent representation amenable to clustering, then fitting lightweight generators within (approximately) unimodal partitions. The objective is cast with mutual-information criteria that link representation quality to sample fidelity. Empirical studies on synthetic discontinuous distributions demonstrate faithful reconstruction and generation.

Self-supervised learning in low-data regimes was addressed by ScatSimCLR framework (Kinakh et al., 2021b) by combining scattering-transform backbones with a pretext regulariser (e.g., predicting transformation parameters). This reduces the number of augmented views and model size while maintaining or improving downstream accuracy. Ablations clarify the roles of the pretext term and view count. Further, the MV–MR approach Kinakh et al. (2024b) advanced self-supervised learning by exploiting multi-views and multi-representations, achieving state-of-the-art performance on STL10 and CIFAR benchmarks while also enabling efficient model-agnostic knowledge distillation.

To address tabular data synthesis, we proposed Binary Diffusion, a generative model based on lossless binary transformations Kinakh and Voloshynovskiy (2024). By leveraging XOR-based noise injection and binary cross-entropy training, this method achieved superior performance on several benchmark datasets while drastically reducing model size.

3.8 Anomaly Detection and High-Energy Physics

Expanding on the program's trajectory from variational autoencoders toward statistically principled template-based pipelines, our work contributed novel frameworks for efficient resonance searches and smooth data morphing. Oleksiyuk et al. (2024b) introduced a model-independent anomaly detection approach using k-means clustering in the space of low-level event observables. By isolating potentially anomalous regions, the method reduced the number of required signal events to achieve a 3σ discovery by nearly 39% compared to conventional parametric background fits, while remaining computationally lightweight for iterative analyses.

Building on these insights, Oleksiyuk et al. (2025) proposed a transport-based adversarial network for smooth conditional mass interpolation, enabling sideband events to be continuously transformed into signal-like distributions. Unlike traditional likelihood-based template generators, TRANSIT achieved competitive anomaly sensitivity with an order-of-magnitude lower training time, making it ideally suited for large-scale iterative scans over multiple mass regions and hypotheses. Its latent representations further provided mass-decorrelated features useful for downstream anomaly searches without introducing mass sculpting. Integrated with existing CURTAINs and DRAPES pipelines, these contributions enrich the program's toolkit for robust, high-sensitivity weakly supervised anomaly detection.

3.9 Machine Learning for Astronomy and Astrophysics

The group made significant advances in bridging machine learning with astronomical imaging, simulation, and inference. Kinakh et al. (2024a) introduced an image-to-image translation framework to predict James Webb Space Telescope (JWST) observations from Hubble Space Telescope (HST) data, comparing Pix2Pix, CycleGAN, TURBO, and Palette DDPM-based models. By incorporating uncertainty quantification within diffusion models, the framework assists astronomers in distinguishing reliable reconstructions from potentially spurious predictions, facilitating efficient planning of JWST observational campaigns.

In radio astronomy, Drozdova et al. (2023) developed a conditional denoising diffusion probabilistic model (DDPM) for reconstructing sky models from interferometric ALMA observations. The model achieved over 90% completeness at low signal-to-noise ratios and significantly outperformed CLEAN and PyBDSF in flux estimation, marking an important milestone for accurate radio-source characterisation. For this contribution, Mariia Drozdova received the SKACH Exceptional PhD Award.

VAEs have been the starting point of this project in solar astronomy. Deep learning approaches are often based on supervised learning and therefore depend on the availability of large, accurately annotated datasets. In practice, obtaining such datasets is often challenging, as it requires extensive human effort and expert labeling. In the context of solar astronomy this challenge is particularly pronounced. Many available datasets are either sparsely labeled, contain ambiguous annotations, or are entirely unlabeled, limiting the applicability of supervised methods. To overcome this data bottleneck, unsupervised deep learning has emerged as a promising alternative, with anomaly detection being one of its most effective applications. We introduced a perspective complementary to the HEP studies, by integrating simulation-based inference and anomaly detection into a unified probabilistic framework (Giger and Csillaghy, 2024). We explored how a purely unsupervised approach can be applied to detect and extract solar phenomena in extreme ultraviolet images from SDO. We demonstrate that a model based on variational autoencoders can effectively identify out-of-distribution samples and localize regions of interest associated with solar activity. By leveraging this unsupervised methodology, we were able to advance automated space weather monitoring tools and contribute to a deeper understanding of the physical processes driving space weather phenomena.

Extending stochastic modelling to solar data, Ramunno et al. (2024a) applied diffusion-based generative models to produce realistic synthetic images of solar flares. They proved that the generated images could be used as a data augmentation technique enhancing the flare prediction results with respect to classical data augmentation. Ramunno et al. (2025) applied latent diffusion models to superresolve the spatial resolution of the Michelson Doppler Imager magnetograms, recovering fine-scale magnetic structures using magnetograms from the Helioseismic Magnetic Imager (HMI) on board SDO. This can be used to have a uniform dataset with the same spatial resolution over two solar cycles starting from 1996 to today. Ramunno et al. (2024b) used denoising diffusion models to forecast the HMI magnetogram 24 hours in advance giving in input the magnetogram at present time as condition. This was one of the first approaches in solar physics of state prediction models, typically used in fluid dynamics, overcoming classical methods such as the persistence model, that consists of assuming that the magnetogram is constant.

Additionally, in collaboration with Polymathic AI, Ramunno prepared a new ML-ready dataset for solar physics, containing all the wavelengths from the SDO telescope. The dataset is presented as a video dataset with a focus on crops of the full disk of the Sun, each video-trajectory contains also a per frame classification in terms flare labels at 1 h, 6h, 12h and 24h from that frame and also time series of the Space Weather HMI Active Region Patch (SHARP) parameters so that it is possible to have a multimodal video dataset with all the physical pre-processing steps taken already in account. The dataset paper is in preparation. The Polymathic AI collaboration brought also to a paper publication at the NeurIPS conference in 2025 about a new inference technique for diffusion models tailored for physics applications with the aim to improve the prediction performances and stabilize the results addressing the exposure bias problem that is well known in long video generations.

In the context of solar flare image reconstruction, Selcuk-Simsek et al. (2025) developed a custom autoencoder, named Fourier Convolutional Decoder (FCD), to retrieve the image of the X-ray emission in solar flares from data recorded by the Spectrometer/Telescope for Imaging X-rays (STIX) on board the ESA Solar Orbiter mission. The model predicts a solar flare image in 0.03 s when run on a CPU, a runtime comparable to the most efficient reconstruction methods. However, FCD outperforms all existing methods in terms of runtime when accelerated by means of GPUs. Since ground truth images are not available for experimental STIX observations, FCD was trained on simulated data. However, the model generalizes well to experimental data as demonstrated by means of an extensive evaluation with several metrics which quantify how well the predicted images reflect the input data measured by the STIX.

A quantized model for predicting the coordinates of flare locations on the solar disk from STIX data was developed by Massa et al. (2024). The model outperforms the algorithm currently implemented onboard STIX in terms of prediction accuracy, and requires a lower number of parameters (21 K vs 34 K). The quantization technique tailored for the developed model allows all the operations to be performed in integer arithmetic without a significant decrease of the model performances.

Massa and Csillaghy (in preparation) developed an unsupervised technique for denoising astronomical images corrupted by Poisson noise by means of a Convolutional Neural Network (CNN). The technique is inspired by the Deep Image Prior approach (DIP) and exploits a recently introduced statistical estimator of the risk for the definition of an appropriate loss function. The method compares favourably to similar approaches that rely on different statistical estimators.

Hackstein et al. (2023) introduced evaluation metrics for galaxy image generators, offering cluster-based metrics and morphology-driven measures that enabled objective, automated quality assessment.

Complementing these advances, Lastufka et al. (2024) benchmarked vision foundation models for optical and radio astronomy, establishing effective fine-tuning strategies to adapt self-supervised and weakly supervised architectures to domain-specific scientific datasets. Self-supervised methods were further extended to wide-field MeerKAT radio images, enabling state-of-the-art morphology classification while drastically reducing preprocessing costs.

Additionally, we investigated the generalisation capabilities of vision-language models (VLMs) for radio galaxy classification Drozdova et al. (2025), revealing both their fragility under prompt variability and their potential when combined with lightweight fine-tuning, achieving up to 97% F1 without domain-specific pretraining.

Finally, motivated by two complementary needs in our generative pipeline: (*i*) safeguarding the provenance of generated images via digital watermarking and (*ii*) quantifying the susceptibility of foundation models to adversarial manipulations, this study establishes a security-centric evaluation of ML watermarking. It formalises attacker models for *copy/transfer* attacks (imprinting a watermark onto clean content) and *removal* attacks (erasing or weakening a watermark while preserving perceptual fidelity), specifies end-to-end threat models and success criteria, and provides a repeatable protocol with calibrated decision thresholds. The empirical analysis spans optimisation-based and latent/FM watermarking schemes, reporting detection trade-offs and failure modes, and releases open artefacts to enable comparable benchmarking (Kinakh et al., 2024c). Thus, we demonstrate that the foundation models are very sensitive to adversarial manipulations and the used protection based digital watermarking is not very robust.

4 Results & outputs

4.1 Anomaly Detection: From Mass Sculpting to Production Deployment

Our anomaly detection research produced a complete methodological transformation in the field. The initial discovery of the "mass sculpting" problem in VAE-based approaches led to the development of the CURTAINs family of methods, culminating in the New Physics Learning Machine (NPLM) that formalizes resonant anomaly detection as density-ratio estimation with improved stability and sensitivity at very low signal fractions.

The practical impact of this work is demonstrated by its deployment within the ATLAS collaboration to search for new physics (Collaboration, 2025), representing a successful transition from research prototype to production analysis tool. This progression from fundamental research to experimental deployment exemplifies the project's success in creating analysis-ready tools for high-energy physics.

4.2 Generative Modeling: Complete Simulation Pipeline

Our generative modeling contributions address the prohibitive computational cost of high-fidelity Monte Carlo simulation. The systematic development from PC-JeDi through EPiC-ly Fast to PC-Droid achieved practical speeds while maintaining high fidelity through improved ODE solvers and distillation techniques. The culmination in PIPPIN - generating full events with variable multiplicities using transformers coupled to diffusion and flow-matching - represents the first practical approach to end-to-end ML-based simulation suitable for High-Luminosity LHC requirements.

These advances directly address one of the field's central needs: scalable, high-fidelity generative simulation for analysis throughput. By progressing from jets to events, from heuristics to likelihood-based methods, this work positions generative modeling as an integral component of the High-Luminosity LHC physics strategy.

4.3 Theoretical Impact and Computational Innovations

The theoretical foundations developed through this project extend well beyond the original physics applications. Our neural sampling methods, free-energy estimation techniques, and flowification framework contribute to the broader machine learning literature while solving specific problems in physics simulations.

Particularly impactful are the computational efficiency innovations. The DeepEMD framework provides two-orders-of-magnitude speedup for point cloud distance computations, enabling practical training of generative models. Our sparse attention mechanisms and Mamba-based architectures address fundamental scalability bottlenecks in large-scale data processing, with applications extending far beyond physics research.

The layer parallelism and outlier feature mitigation techniques developed for transformer architectures have broad applicability in the current landscape of large language models and foundation model research.

4.4 Cross-Domain Validation and Community Impact

The successful adaptation of our methods to astrophysical data through SkyCURTAINs provides crucial validation of method robustness and generalizability. Recovering stellar streams such as GD-1 with higher purity than previous approaches in billion-object astronomical surveys demonstrates that the principles developed for collider physics have broader scientific applicability.

Our contributions to foundation models in high energy physics, particularly the Masked Particle Model and its improvements, have established new paradigms for machine learning applications in physics, with demonstrated improvements in dataset size efficiency and training time across many tasks and datasets.

4.5 Publications and Software Outputs

The project has resulted in over 25 peer-reviewed publications in top-tier venues including multiple NeurIPS conferences, with strong impact across both machine learning and physics communities. All major methodological contributions have been released as open-source software with comprehensive documentation, enabling community adoption and further development.

Software packages include complete implementations of the CURTAINs framework, NPLM, all generative models (PC-JeDi, EPiC-ly Fast, PC-Droid, PIPPIN), DeepEMD, sparse attention mechanisms, and Mamba-based architectures. The SUPA simulator provides new benchmark datasets for calorimeter simulation research with up to 6 orders of magnitude speedup over Geant4, representing a significant community resource.

The Fourier Convolutional Decoder (FCD) model has been released as (i) a full open-source training code repository on GitHub[1], (ii) a pretrained model[2], and (iii) an interactive demonstration on Hugging Face[3].

5 Deviations

There were no deviations from the approved research plan or milestones. All deliverables were completed according to schedule. No technical, ethical, or data-related risks were encountered during the project.

[1] https://github.com/i4Ds/fcd
[2] https://huggingface.co/mervess/FCD-Solar
[3] https://huggingface.co/spaces/mervess/FCD-Solar-Demo

References

G. Aad, B. Abbott, J. Abdallah, A. Abdelalim, A. Abdesselam, B. Abi, M. Abolins, H. Abramowicz, H. Abreu, B. Acharya, et al. The atlas simulation infrastructure. *The European Physical Journal C*, 70(3):823–874, 2010.

G. Aad et al. A continuous calibration of the ATLAS flavour-tagging classifiers via optimal transportation maps. 5 2025.

S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8. URL https://www.sciencedirect.com/science/article/pii/S0168900203013688.

M. Algren, T. Golling, M. Guth, C. Pollard, and J. A. Raine. Flow Away your Differences: Conditional Normalizing Flows as an Improvement to Reweighting. 4 2023.

M. Algren, J. A. Raine, and T. Golling. Decorrelation using optimal transport. *Eur. Phys. J. C*, 84(6):579, 2024. doi: 10.1140/epjc/s10052-024-12868-6.

M. Algren, T. Golling, F. A. Di Bello, and C. Pollard. Mind the Gap: Navigating Inference with Optimal Transport Maps. 7 2025a.

M. Algren, T. Golling, C. Pollard, and J. A. Raine. Variational inference for pile-up removal at hadron colliders with diffusion models. *Phys. Rev. D*, 111(11):116010, 2025b. doi: 10.1103/PhysRevD.111.116010.

K. G. Barman et al. Large physics models: towards a collaborative approach with large language models and foundation models. *Eur. Phys. J. C*, 85(9):1066, 2025. doi: 10.1140/epjc/s10052-025-14707-8.

D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan. Sampling using SU(n) gauge equivariant flows. *Phys. Rev. D*, 103:074504, Apr 2021. doi: 10.1103/PhysRevD.103.074504. URL https://link.aps.org/doi/10.1103/PhysRevD.103.074504.

E. Buhmann, C. Ewen, D. A. Faroughy, T. Golling, G. Kasieczka, M. Leigh, G. Quétant, J. A. Raine, D. Sengupta, and D. Shih. Epic-ly fast particle cloud generation with flow-matching and diffusion. *arXiv preprint arXiv:2310.00049*, 2023.

S. Caron et al. Strategic White Paper on AI Infrastructure for Particle, Nuclear, and Astroparticle Physics: Insights from JENA and EuCAIF. 3 2025.

A. Collaboration. Weakly supervised anomaly detection for resonant new physics in the dijet final state using proton-proton collisions at $\sqrt{s} = 13$ tev with the atlas detector, 2025. URL https://arxiv.org/abs/2502.09770.

M. Drozdova, V. Kinakh, G. Quétant, T. Golling, and S. Voloshynovskiy. Generation of data on discontinuous manifolds via continuous stochastic non-invertible networks. *NeurIPS 2021 Workshop on Bayesian Deep Learning*, 2021. Poster presentation at NeurIPS 2021 Bayesian Deep Learning.

M. Drozdova, V. Kinakh, O. Bait, O. Taran, E. Lastufka, M. Dessauges-Zavadsky, T. Holotyak, D. Schaerer, and S. Voloshynovskiy. Radio-astronomical image reconstruction with a conditional denoising diffusion model. *Astronomy & Astrophysics*, 12 2023. doi: 10.1051/0004-6361/202347948.

M. Drozdova, E. Lastufka, V. Kinakh, T. Holotyak, D. Schaerer, and S. Voloshynovskiy. Radio astronomy in the era of vision-language models: Prompt sensitivity and adaptation. *NeurIPS 2025 Workshop on Machine Learning and the Physical Sciences*, 2025.

V. P. Dwivedi and X. Bresson. A generalization of transformer networks to graphs. *CoRR*, abs/2012.09699, 2020. URL https://arxiv.org/abs/2012.09699.

M. Giger and A. Csillaghy. Unsupervised anomaly detection with variational autoencoders applied to full-disk solar images. *Space Weather*, 22(2):e2023SW003516, 2024. doi: https://doi.org/10.1029/2023SW003516. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023SW003516. e2023SW003516 2023SW003516.

T. Golling, S. Klein, R. Mastandrea, and B. Nachman. Flow-enhanced transportation for anomaly detection. *Phys. Rev. D*, 107(9):096025, 2023a. doi: 10.1103/PhysRevD.107.096025.

T. Golling, S. Klein, R. Mastandrea, B. Nachman, and J. A. Raine. Morphing one dataset into another with maximum likelihood estimation. *Phys. Rev. D*, 108(9):096018, 2023b. doi: 10.1103/PhysRevD.108.096018.

T. Golling, T. Nobe, D. Proios, J. A. Raine, D. Sengupta, S. Voloshynovskiy, J.-F. Arguin, J. L. Martin, J. Pilette, D. B. Gupta, and A. Farbin. The mass-ive issue: Anomaly detection in jet physics, 2023c. URL https://arxiv.org/abs/2303.14134.

T. Golling, L. Heinrich, M. Kagan, S. Klein, M. Leigh, M. Osadchy, and J. Andrew Raine. Masked particle modeling on sets: towards self-supervised high energy physics foundation models. *Machine Learning: Science and Technology*, 5(3):035074, sep 2024a. doi: 10.1088/2632-2153/ad64a8. URL https://doi.org/10.1088/2632-2153/ad64a8.

T. Golling, G. Kasieczka, C. Krause, R. Mastandrea, B. Nachman, J. A. Raine, D. Sengupta, D. Shih, and M. Sommerhalder. The interplay of machine learning-based resonant anomaly detection methods. *Eur. Phys. J. C*, 84(3):241, 2024b. doi: 10.1140/epjc/s10052-024-12607-x.

R. C. González, D. Paliotta, M. Pagliardini, M. Jaggi, and F. Fleuret. Leveraging the true depth of llms. In *ICLR 2025 Workshop on Foundation Models in the Wild*, 2025.

G. Grosso, D. Sengupta, T. Golling, and P. Harris. Robust resonant anomaly detection with nplm. *The European Physical Journal C*, 85(9):1074, 2025.

S. Hackstein, V. Kinakh, C. Bailer, and M. Melchior. Evaluation metrics for galaxy image generators. *Astronomy and Computing*, 42:100685, 2023.

B. He, L. Noci, D. Paliotta, I. Schlag, and T. Hofmann. Understanding and minimising outlier features in neural network training. In *ICML 2024 Workshop on Theoretical Foundations of Foundation Models*, 2024.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022. URL https://arxiv.org/abs/2212.13345.

C.-W. Huang, L. Dinh, and A. Courville. Augmented normalizing flows: Bridging the gap between generative flows and latent variable models. *arXiv preprint arXiv:2002.07101*, 2020.

V. Kinakh and S. Voloshynovskiy. Tabular data generation using binary diffusion. *arXiv preprint arXiv:2409.13882*, 2024. URL https://github.com/vkinakh/binary-diffusion-tabular. NeurIPS 2024 Workshop on Table Representation Learning.

V. Kinakh, M. Drozdova, G. Quétant, T. Golling, and S. Voloshynovskiy. Information-theoretic stochastic contrastive conditional gan: Infoscc-gan. *arXiv preprint arXiv:2112.09653*, 2021a. URL https://github.com/vkinakh/InfoSCC-GAN. NeurIPS 2021 Workshop on Bayesian Deep Learning.

V. Kinakh, O. Taran, and S. Voloshynovskiy. Scatsimclr: self-supervised contrastive learning with pretext task regularization for small-scale datasets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1098–1106, 2021b. URL https://vipriors.github.io/.

V. Kinakh, Y. Belousov, G. Quétant, M. Drozdova, T. Holotyak, D. Schaerer, and S. Voloshynovskiy. Hubble meets webb: Image-to-image translation in astronomy. *Sensors*, 24(4):1151, 2024a.

V. Kinakh, M. Drozdova, and S. Voloshynovskiy. Mv–mr: Multi-views and multi-representations for self-supervised learning and knowledge distillation. *Entropy*, 26(6):466, 2024b.

V. Kinakh, B. Pulfer, Y. Belousov, P. Fernandez, T. Furon, and S. Voloshynovskiy. Evaluation of security of ml-based watermarking: Copy and removal attacks. In *2024 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2024c. URL https://github.com/vkinakh/ssl-watermarking-attacks.

S. Klein and T. Golling. Decorrelation with conditional normalizing flows. 11 2022.

E. Lastufka, O. Bait, O. Taran, M. Drozdova, V. Kinakh, D. Piras, M. Audard, M. Dessauges-Zavadsky, T. Holotyak, D. Schaerer, et al. Self-supervised learning on meerkat wide-field continuum images. *Astronomy & Astrophysics*, 690:A310, 2024.

M. Leigh, J. A. Raine, K. Zoch, and T. Golling. ν-flows: Conditional neutrino regression. *SciPost Phys.*, 14(6):159, 2023. doi: 10.21468/SciPostPhys.14.6.159.

M. Leigh, D. Sengupta, B. Nachman, and T. Golling. Accelerating template generation in resonant anomaly detection searches with optimal transport. 7 2024a.

M. Leigh, D. Sengupta, G. Quétant, J. A. Raine, K. Zoch, and T. Golling. Pc-jedi: Diffusion for particle cloud generation in high energy physics. *SciPost Physics*, 16(1):018, 2024b.

M. Leigh, D. Sengupta, J. A. Raine, G. Quétant, and T. Golling. Faster diffusion model with improved quality for particle cloud generation. *Physical Review D*, 109(1):012010, 2024c.

M. Leigh, S. Klein, F. Charton, T. Golling, L. Heinrich, M. Kagan, I. Ochoa, and M. Osadchy. Is tokenization needed for masked particle modeling? *Machine Learning: Science and Technology*, 6(2):025075, jun 2025.

P. Massa, S. Felix, L. I. Etesi, E. C. M. Dickson, H. Xiao, F. P. Ramunno, M. Selcuk-Simsek, B. Panos, A. Csillaghy, and S. Krucker. A machine learning approach for computing solar flare locations in x-rays on-board solar orbiter/stix, Oct. 2024. URL https://doi.org/10.5281/zenodo.13885681.

B. Máté and F. Fleuret. Deformations of boltzmann distributions. In *Proceedings of the NeurIPS workshop on Machine Learning and the Physical Sciences*, 2022.

B. Máté and F. Fleuret. Learning interpolations between boltzmann densities. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=TH6YrEcbth.

B. Mate and F. Fleuret. Multi-lattice sampling of quantum field theories via neural operator-based flows. *Machine Learning: Science and Technology*, 5(4):045053, 2024.

B. Máté, S. Klein, T. Golling, and F. Fleuret. Flowification: Everything is a normalizing flow. *Advances in Neural Information Processing Systems*, 35:35478–35489, 2022.

B. Mate, F. Fleuret, and T. Bereau. Neural thermodynamic integration: Free energies from energy-based diffusion models. *The Journal of Physical Chemistry Letters*, 15(45):11395–11404, 2024.

B. Máté, F. Fleuret, and T. Bereau. Solvation free energies from neural thermodynamic integration. *The Journal of Chemical Physics*, 162(12):124107, 03 2025. ISSN 0021-9606. doi: 10.1063/5.0251736. URL https://doi.org/10.1063/5.0251736.

D. Nielsen, P. Jaini, E. Hoogeboom, O. Winther, and M. Welling. Survae flows: Surjections to bridge the gap between vaes and flows. *Advances in Neural Information Processing Systems*, 33:12685–12696, 2020.

F. Noé, S. Olsson, J. Köhler, and H. Wu. Boltzmann generators – sampling equilibrium states of many-body systems with deep learning, 2018. URL https://arxiv.org/abs/1812.01729.

I. Oleksiyuk, J. A. Raine, M. Krämer, S. Voloshynovskiy, and T. Golling. Cluster Scanning: a novel approach to resonance searches. *JHEP*, 06:163, 2024a. doi: 10.1007/JHEP06(2024)163.

I. Oleksiyuk, J. A. Raine, M. Krämer, S. Voloshynovskiy, and T. Golling. Cluster scanning: a novel approach to resonance searches. *Journal of High Energy Physics*, 2024(6):1–33, 2024b.

I. Oleksiyuk, S. Voloshynovskiy, and T. Golling. Transit your events into a new mass: Fast background interpolation for weakly-supervised anomaly searches. *JHEP 07 (2025) 177*, 2025.

M. Pagliardini, D. Paliotta, M. Jaggi, and F. Fleuret. Fast attention over long sequences with dynamic sparse flash attention. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc. URL https://arxiv.org/abs/2306.01160.

D. Paliotta, J. Wang, M. Pagliardini, K. Y. Li, A. Bick, J. Z. Kolter, A. Gu, F. Fleuret, and T. Dao. Thinking slow, fast: Scaling inference compute with distilled reasoners, 2025. URL https://arxiv.org/abs/2502.20339.

G. Quetant, M. Drozdova, V. Kinakh, T. Golling, and S. Voloshynovskiy. Turbo-sim: a generalised generative model with a physical latent space, 2021. URL https://arxiv.org/abs/2112.10629.

G. Quétant, Y. Belousov, V. Kinakh, and S. Voloshynovskiy. Turbo: The swiss knife of auto-encoders. *Entropy*, 25(10):1471, 2023.

G. Quétant, J. A. Raine, M. Leigh, D. Sengupta, and T. Golling. Generating variable length full events from partons. *Physical Review D*, 110(7):076023, 2024.

J. A. Raine, S. Klein, D. Sengupta, and T. Golling. Curtains for your sliding window: Constructing unobserved regions by transforming adjacent intervals. *Frontiers in Big Data*, 6, Mar. 2023. ISSN 2624-909X. doi: 10.3389/fdata.2023.899345. URL http://dx.doi.org/10.3389/fdata.2023.899345.

J. A. Raine, M. Leigh, K. Zoch, and T. Golling. Fast and improved neutrino reconstruction in multineutrino final states with conditional normalizing flows. *Phys. Rev. D*, 109(1):012005, 2024. doi: 10.1103/PhysRevD.109.012005.

F. P. Ramunno, S. Hackstein, V. Kinakh, M. Drozdova, G. Quétant, A. Csillaghy, and S. Voloshynovskiy. Solar synthetic imaging: Introducing denoising diffusion probabilistic models on sdo/aia data. *Astronomy & Astrophysics*, 686:A285, 2024a.

F. P. Ramunno, H.-J. Jeong, S. Hackstein, A. Csillaghy, S. Voloshynovskiy, and M. K. Georgoulis. Magnetogram-to-magnetogram: Generative forecasting of solar evolution, Oct. 2024b. URL https://doi.org/10.5281/zenodo.13885515.

F. P. Ramunno, P. Massa, V. Kinakh, B. Panos, A. Csillaghy, and S. Voloshynovskiy. Enhancing image resolution of solar magnetograms: A latent diffusion model approach. *Astronomy & Astrophysics*, 698:A140, 2025.

F. Rothen, S. Klein, M. Leigh, and T. Golling. Enhancing generalization in high-energy physics using white-box adversarial attacks. *Phys. Rev. D*, 112(1):016004, 2025. doi: 10.1103/PhysRevD.112.016004.

M. Selcuk-Simsek, P. Massa, H. Xiao, S. Krucker, and A. Csillaghy. Fourier convolutional decoder: reconstructing solar flare images via deep learning. *Neural Computing and Applications*, pages 1–32, 2025.

D. Sengupta, S. Klein, J. A. Raine, and T. Golling. Curtains flows for flows: Constructing unobserved regions with maximum likelihood estimation. *SciPost Physics*, 17(2):046, 2024a.

D. Sengupta, M. Leigh, J. A. Raine, S. Klein, and T. Golling. Improving new physics searches with diffusion models for event observables and jet constituents. *JHEP*, 04:109, 2024b. doi: 10.1007/JHEP04(2024)109.

D. Sengupta, M. Leigh, J. A. Raine, S. Klein, and T. Golling. Improving new physics searches with diffusion models for event observables and jet constituents. *Journal of High Energy Physics*, 2024(4):1–32, 2024c.

D. Sengupta, S. Mulligan, D. Shih, J. A. Raine, and T. Golling. Skycurtains: model-agnostic search for stellar streams with gaia data. *Monthly Notices of the Royal Astronomical Society*, 536(2):1104–1114, 2025.

A. K. Sinha and F. Fleuret. Deepemd: A transformer-based fast estimation of the earth mover's distance. In *International Conference on Pattern Recognition*, pages 1–15. Springer, 2024.

A. K. Sinha, D. Paliotta, B. Máté, J. Raine, T. Golling, and F. Fleuret. Supa: A lightweight diagnostic simulator for machine learning in particle physics. *Advances in Neural Information Processing Systems*, 36:64829–64856, 2023.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

J. Wang, D. Paliotta, A. May, A. Rush, and T. Dao. The mamba in the llama: Distilling and accelerating hybrid models. *Advances in Neural Information Processing Systems*, 37:62432–62457, 2024.

J. Wang, W.-D. Li, D. Paliotta, D. Ritter, A. M. Rush, and T. Dao. M1: Towards scalable test-time compute with mamba reasoning models, 2025. URL https://arxiv.org/abs/2504.10449.