Infiltration of AI/ML in Particle Physics

Aspects of data handling

Snowmass Community Planning Meeting
Oct 5-8, 2020

Jean-Roch Vlimant
jvlimant@caltech.edu @vlimant

Outline

I. Motivations for Machine Learning

II. Path to the Future of AI in HEP

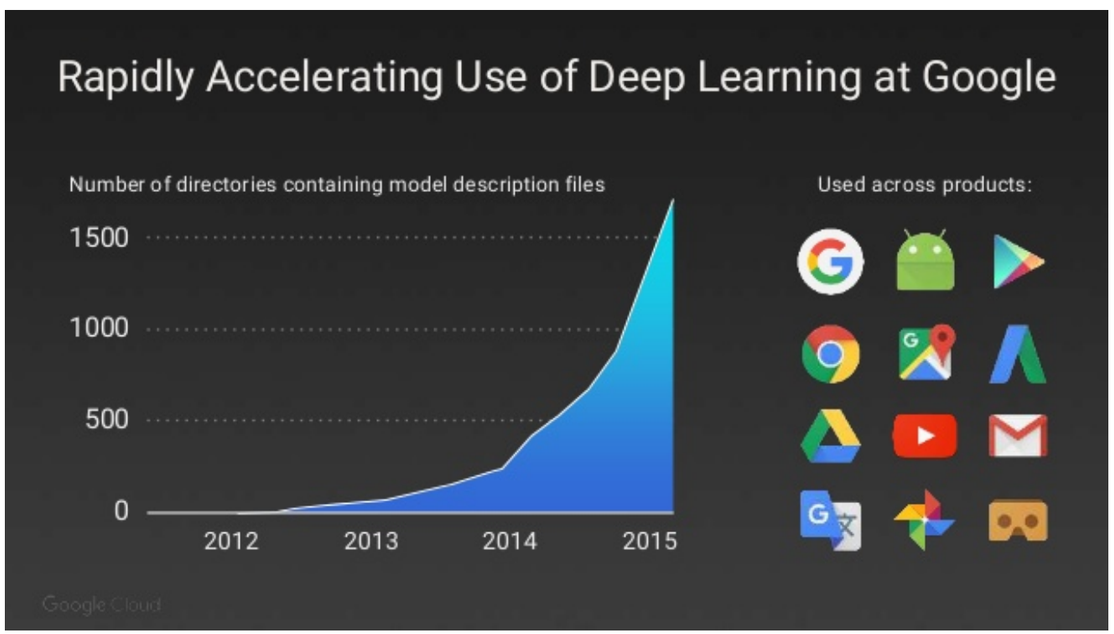
Motivations for Using Machine Learning in High Energy Physics

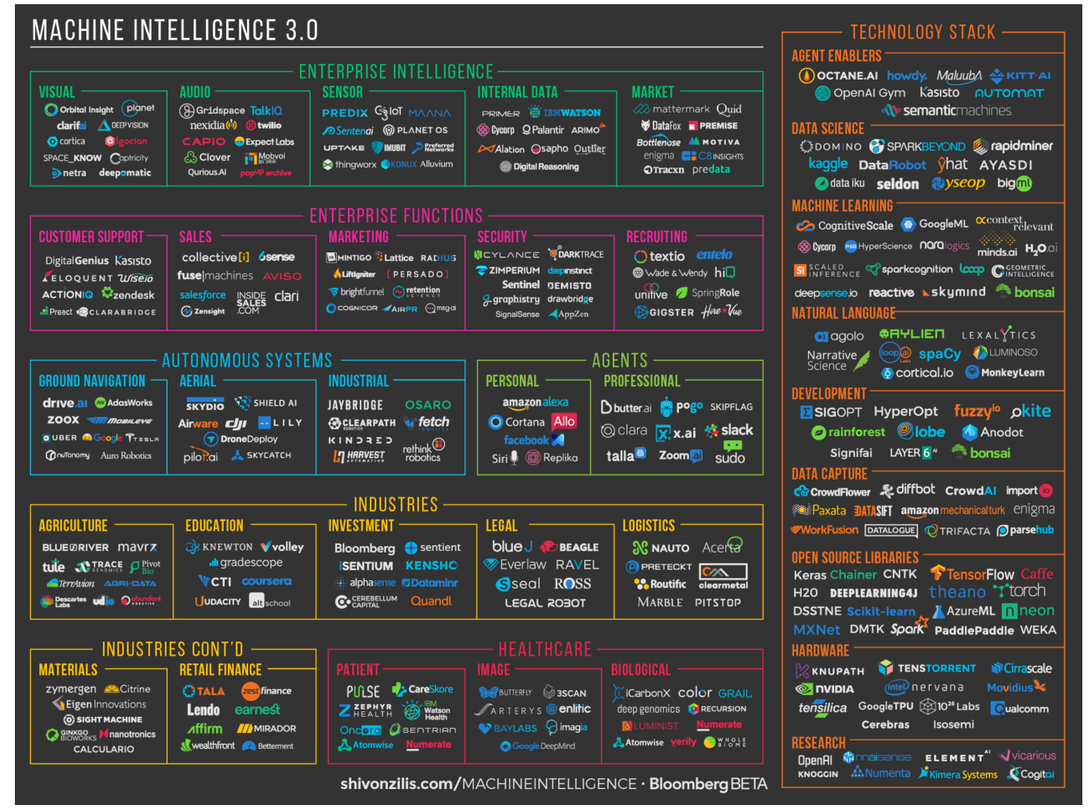
and elsewhere ...

Gerd

Machine Learning in Industry

https://www.nvidia.com/en-us/deep-learning-ai/

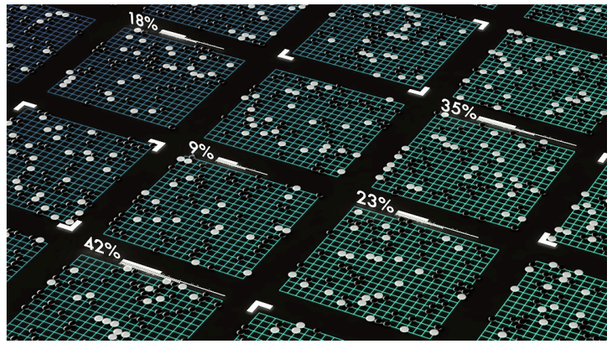


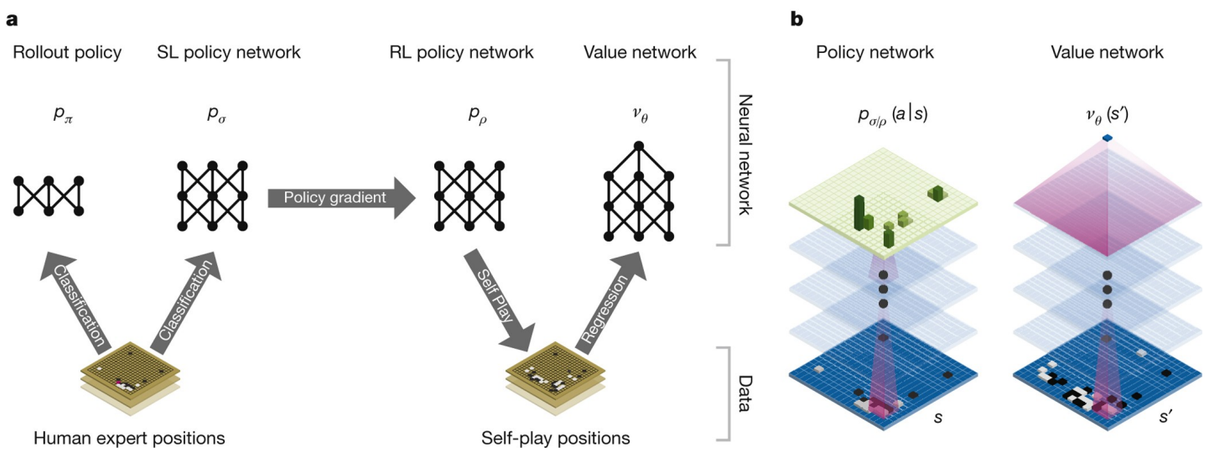


http://www.shivonzilis.com/machineintelligence

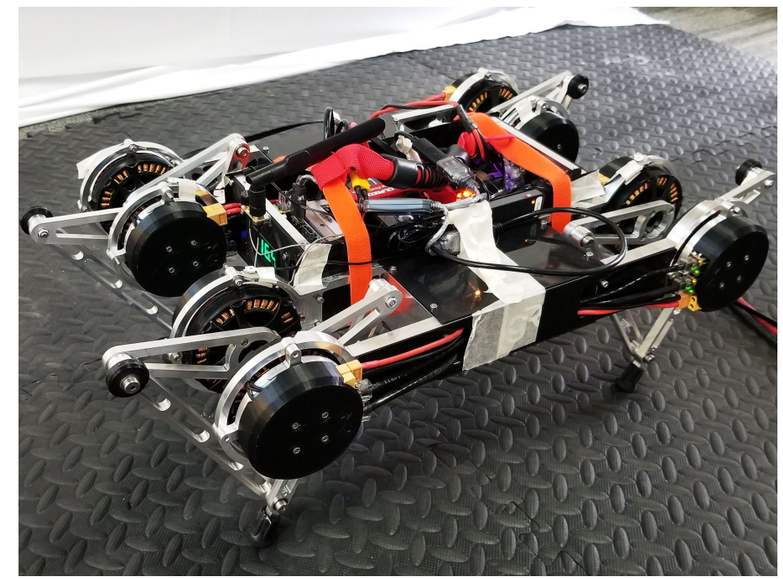
- Prominent field in industry nowadays
- Lots of data, lots of applications, lots of potential use cases, lots of money
- → Knowing machine learning can open significantly **career horizons**

Learning to Control





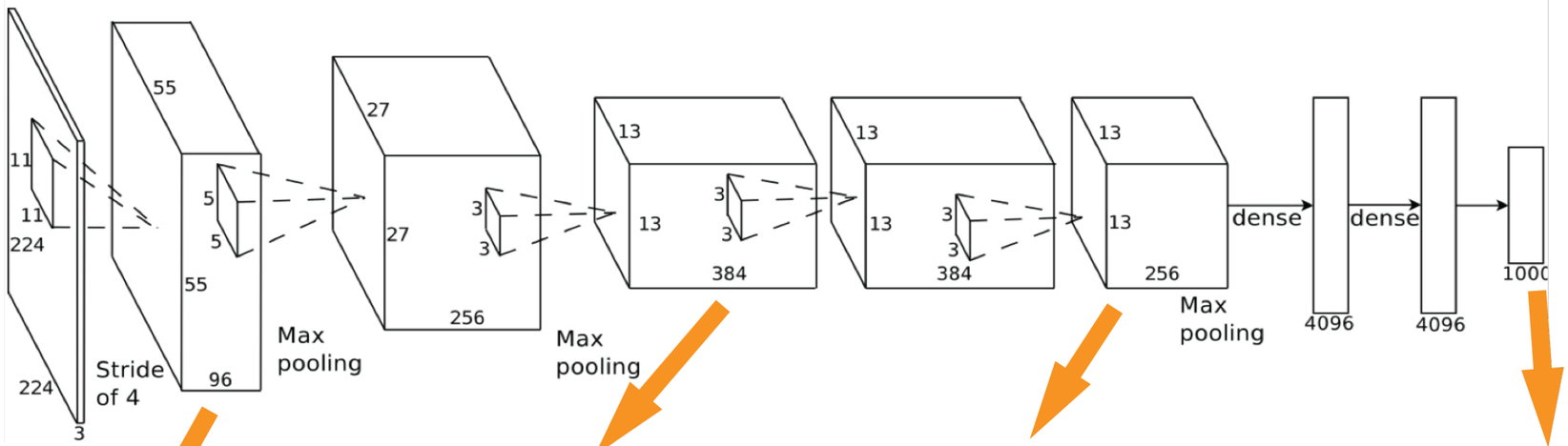
Mastering the game of Go with deep neural networks and tree search,
https://doi.org/10.1038/nature16961

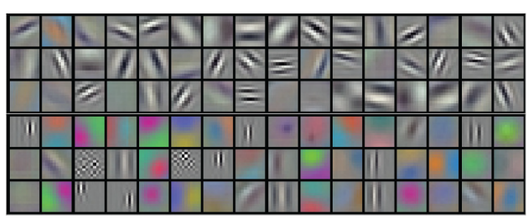


Learning to Walk via Deep Reinforcement Learning
https://arxiv.org/abs/1812.11103

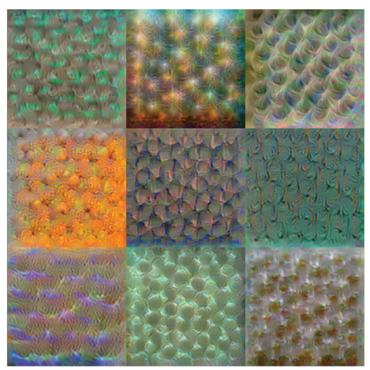
Modern machine learning **boosts control technologies**.
AI, gaming, robotic, self-driving vehicle, etc.

Learning from Complexity

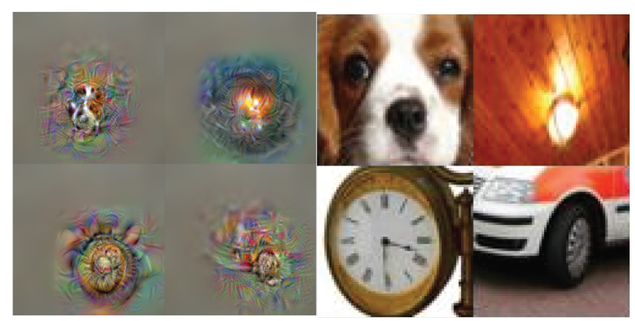




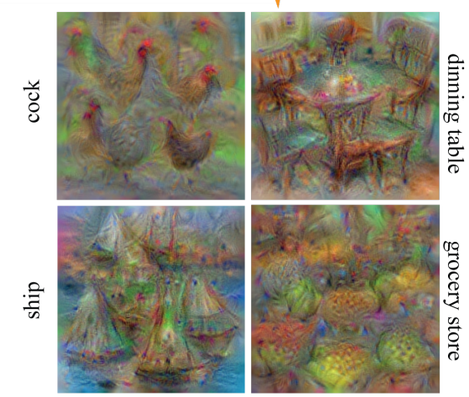
Conv 1: Edge+Blob



Conv 3: Texture



Conv 5: Object Parts

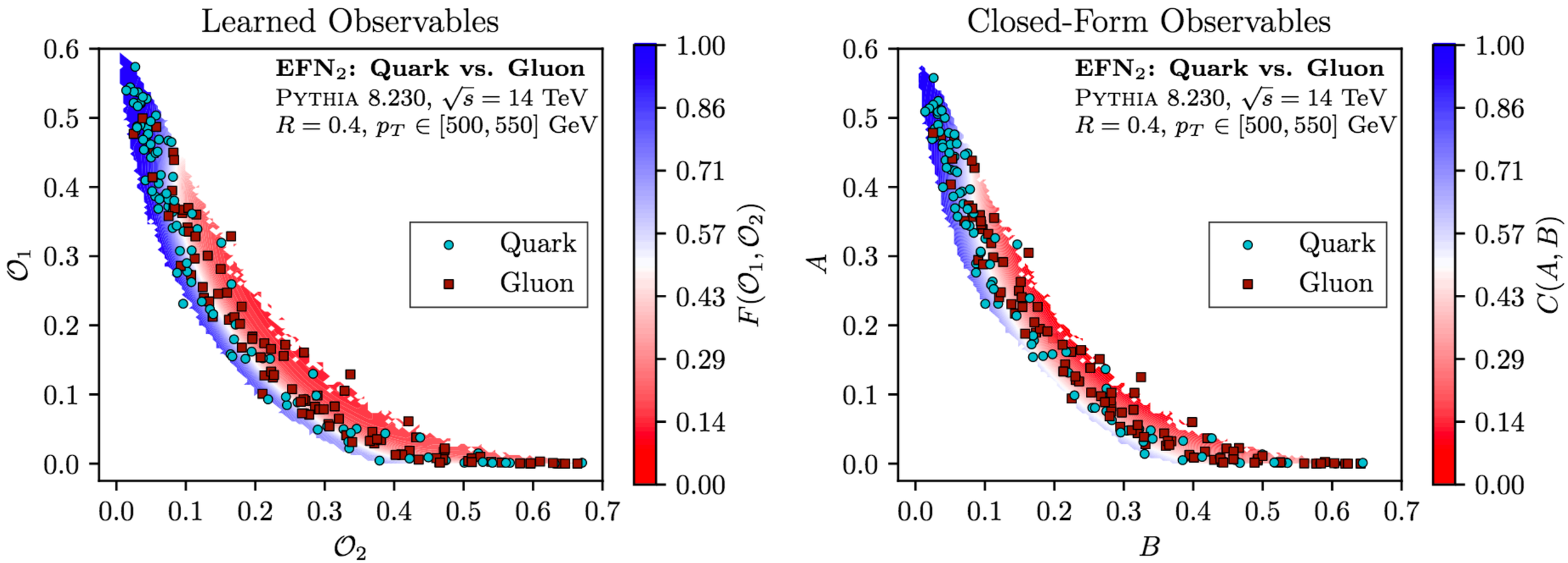


Fc8: Object Classes

Machine learning model can **extract information from complex dataset.**

More classical algorithm counter part may take **years of development.**

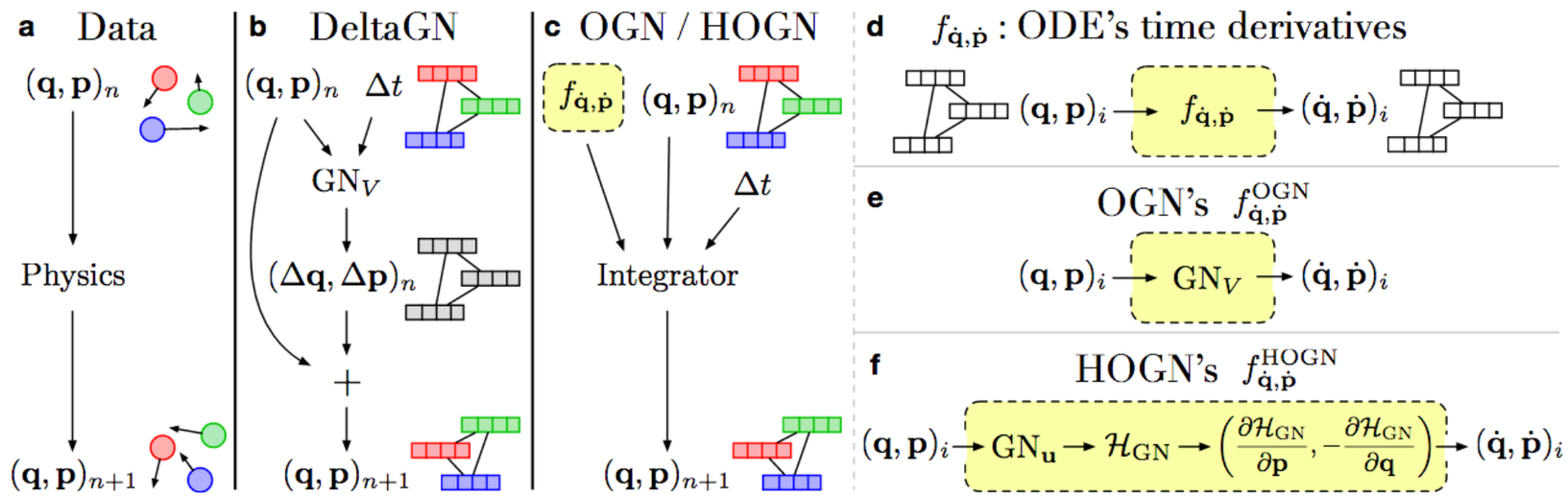
Learn Physics

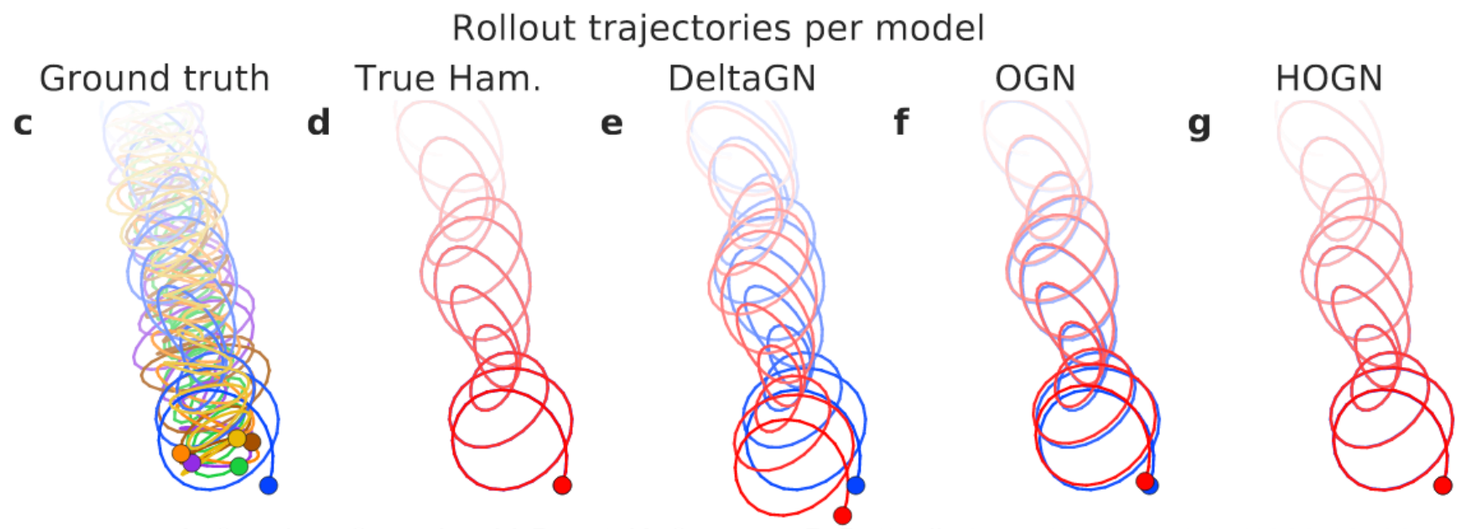


P. Komiske, E. Metodiev, J. Thaler, https://arxiv.org/abs/1810.05165

Machine Learning can **help understand Physics**.

Use Physics

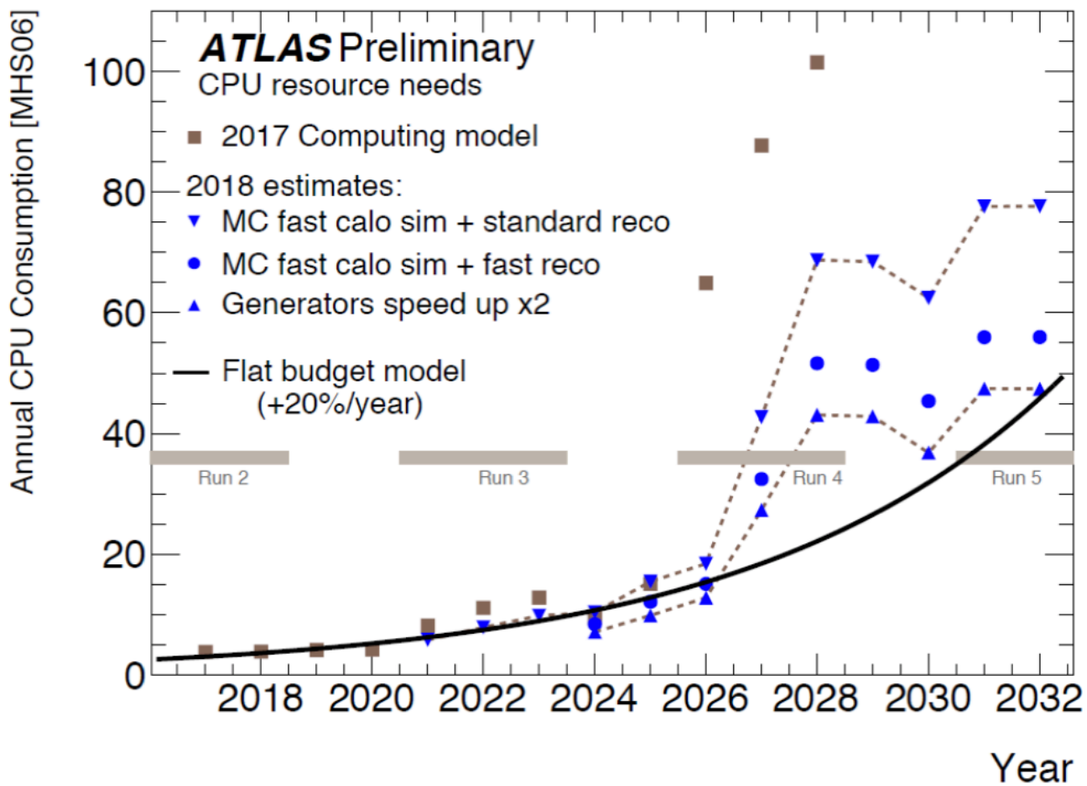


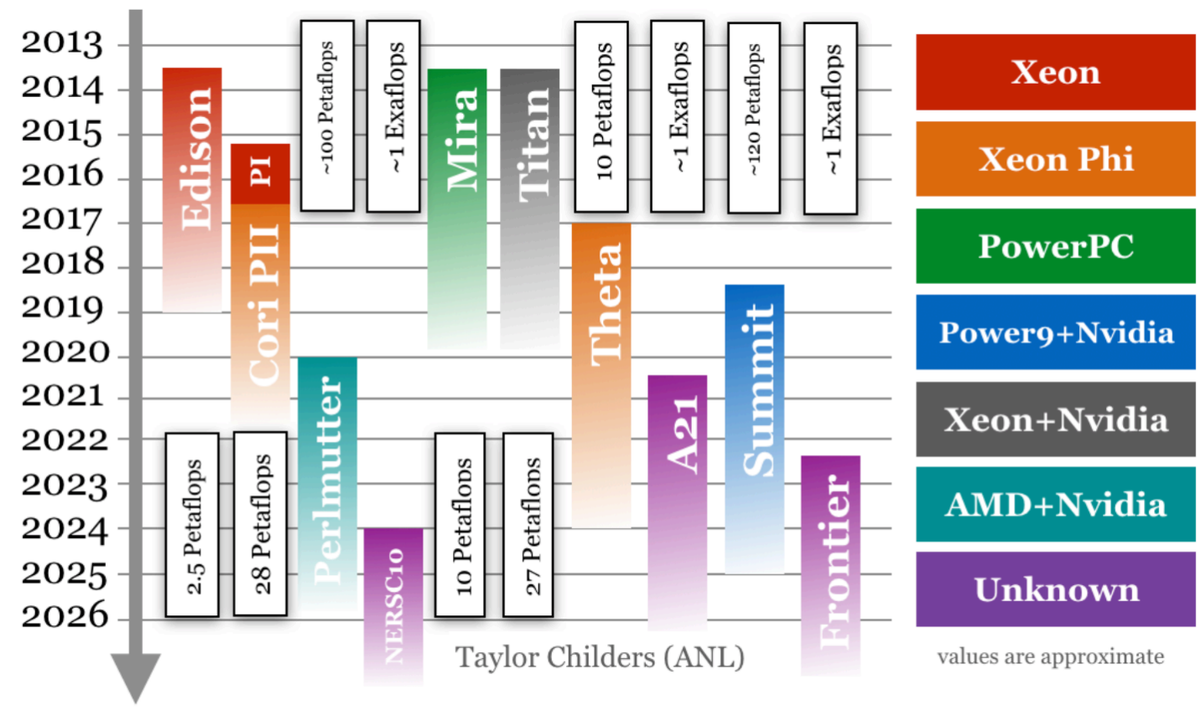


A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, P. Battaglia https://arxiv.org/abs/1909.12790

Let me model **include Physics principles** to master convergence

The Computing Cost of Science

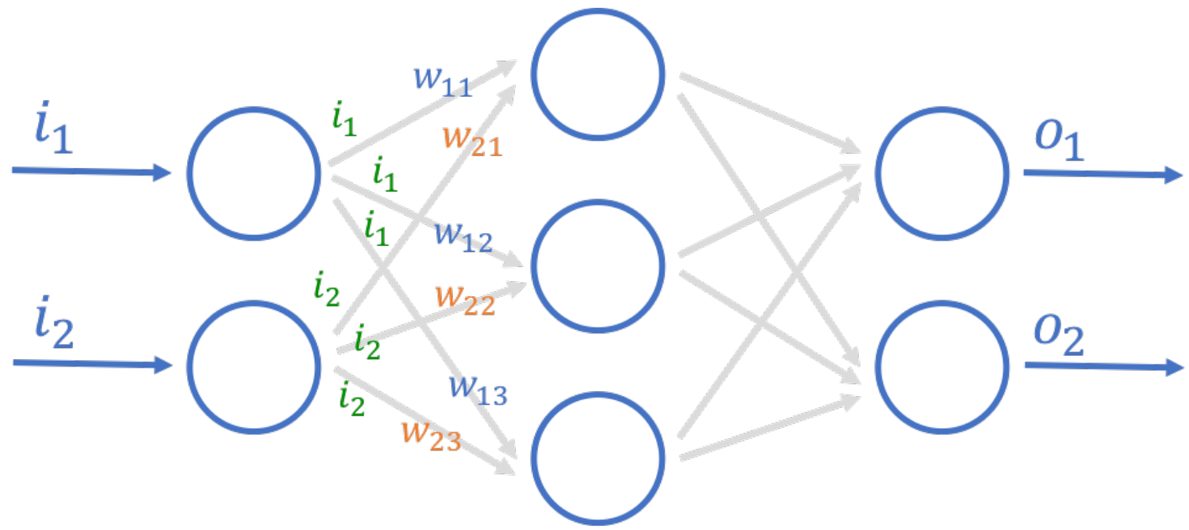




https://indico.cern.ch/event/822126/contributions/3500169/

Ever growing needs for computing resource
Slowdown of classical architecture, growth of GPU architecture

Operation Vectorization

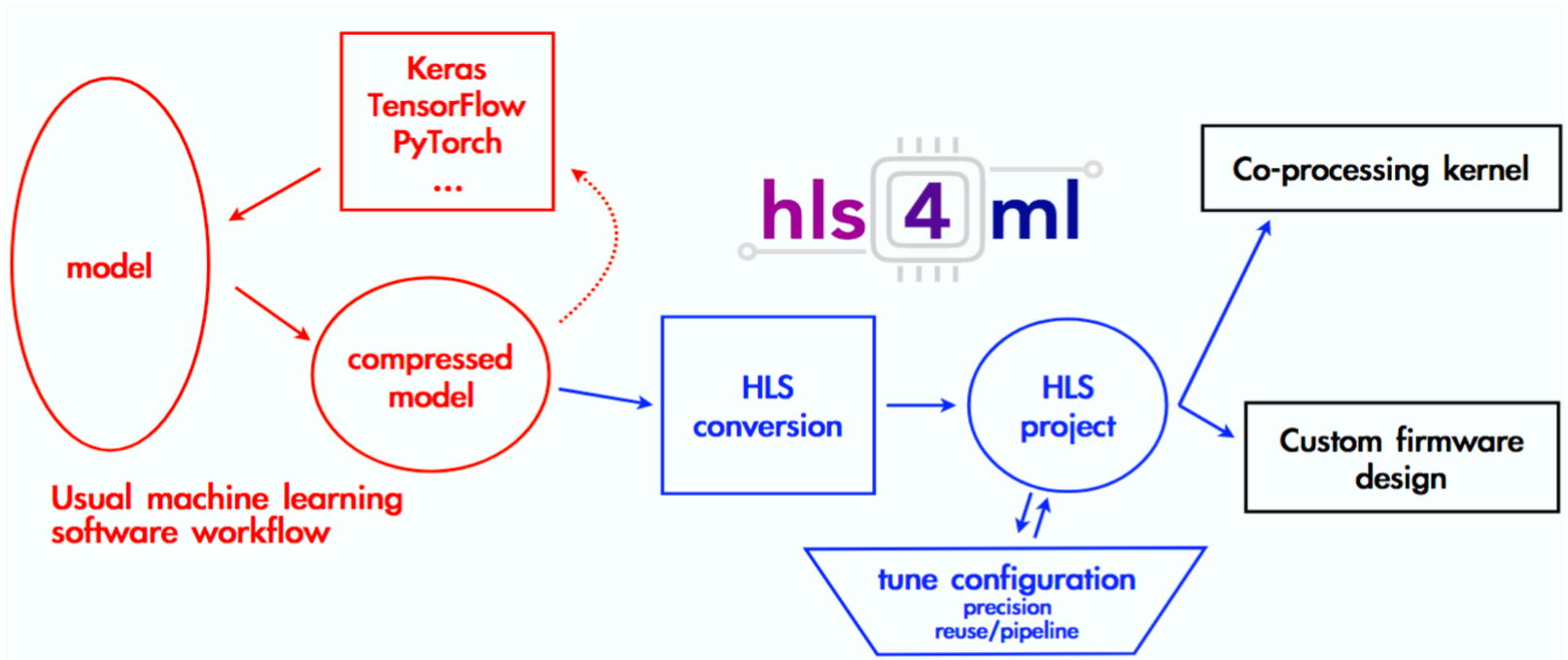


ANN ≡ matrix operations ≡ parallelizable

$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix}$$

Computation of prediction from artificial neural network model can be **vectorized to a large extend.**

Hyper-Fast Prediction



Synthesizing FPGA firmware from trained ANN

https://hls-fpga-machine-learning.github.io/hls4ml/

J. Duarte et al.https://arxiv.org/abs/1804.06913

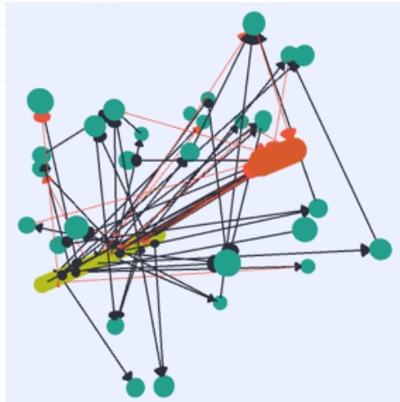
Prediction from artificial neural network model can be **done on FPGA**, GPU, TPU, ...

Low Power Prediction

Best Results: Single View

x-view (127x50)	conv1 (8x3)	pool1 (2x1)	conv2 (7x3)	pool2 (2x1)	conv3 (6x3)	pool3 (2x1)	conv4 (6x3)	pool4 (2x1)	fc1 (196)	drop out	fc2 (98)	drop out	fc3 (11)	classification

Convolutional Neural Network Result: ~80.42%



- 90 neurons, 86 synapses
- Estimated energy for a single classification for mrDANNA implementation: 1.66 μJ

Spiking Neural Network Result: ~80.63%

Source for CNN results: A. Terwilliger, et al. Vertex Reconstruction of Neutrino Interactions using Deep Learning. IJCNN 2017.

33 Programming Neuromorphic Computing Systems

OAK RIDGE National Laboratory

https://indico.fnal.gov/event/13497/contribution/0 *Slide C. Schuman*

Neuromorphic hardware dedicated to spiking neural networks.
Low power consumption by design.

The Future of AI in High Energy Physics

as I see it ...

Gerd

Unrolling Success Stories

Machine learning already deployed in many areas of HEP data analysis

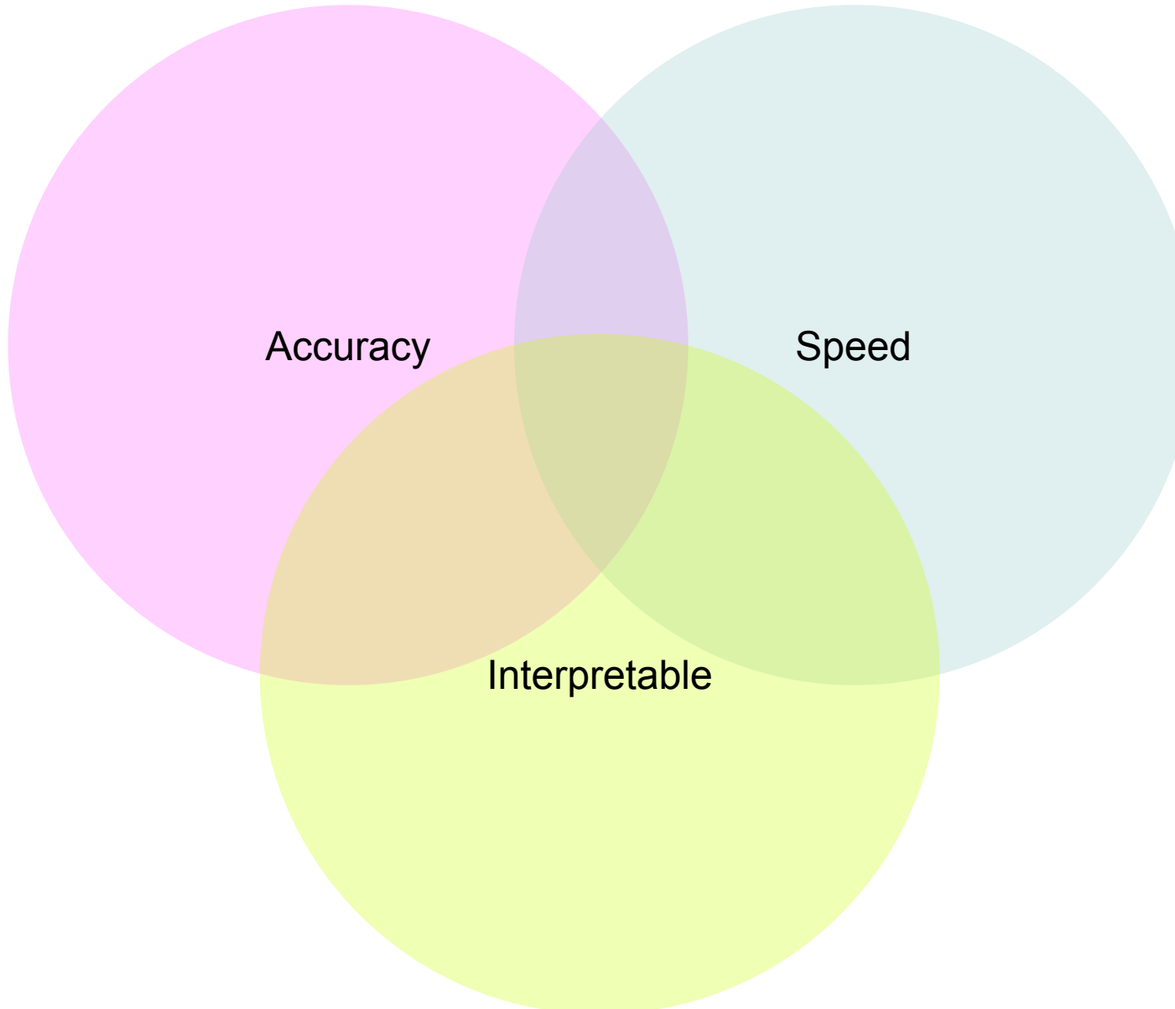
- Trigger decisions
- Signal classification
- Signal region categorization
- Jet tagging, particle identification, classification within reconstruction
- Energy regression
- …

"Classical" MVA (Boosted decision trees, support vector machine, artificial neural network) mostly used.
Increasing usage of deeper/complex artificial neural network (CNN, RNN, …)

Deep learning R&D to go beyond the performance and these type applications

- ➡Generative models
- ➡Reconstruction
- ➡Anomaly detection
- ➡Theory Inference
- ➡Control technology
- ➡…

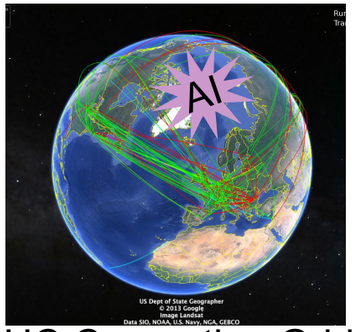
Possible Utilizations



- **Fast surrogate** models (trigger, simulation, etc) ; even better if more accurate.
- **More accurate** than existing algorithms (tagging, regression, etc) ; even better if faster.
- Model performing **otherwise impossible tasks** (operations, etc)

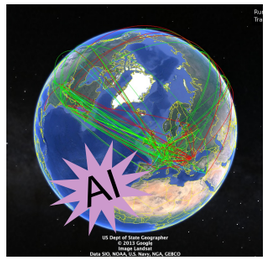
AI in HEP

Role of AI: accelerator control, data acquisition, event triggering, anomaly detection, new physics scouting, event reconstruction, event generation, detector simulation, LHC grid control, analytics, signal extraction, likelihood free inference, background rejection, new physics searches, ...

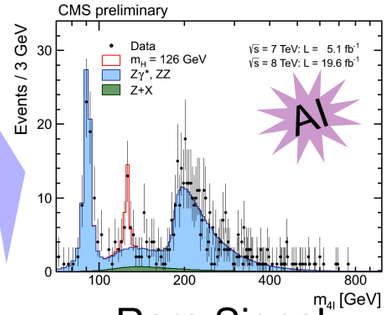


LHC Computing Grid
200k cores pledge to
CMS over ~100 sites

CERN Tier-0/Tier-1
Tape Storage
200PB total

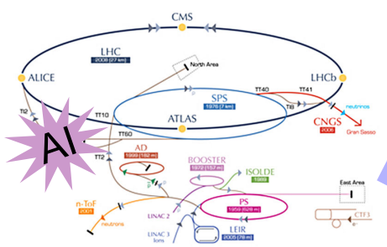


LHC Grid
Remote Access
to 100PB of data



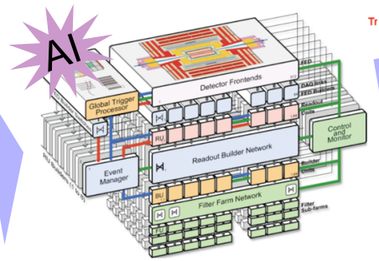
Rare Signal
Measurement
~1 out of 10^6

CERN Tier-0
Computing Center
20k cores



Large Hadron Collider
40 MHz of collision

CMS Detector
1PB/s



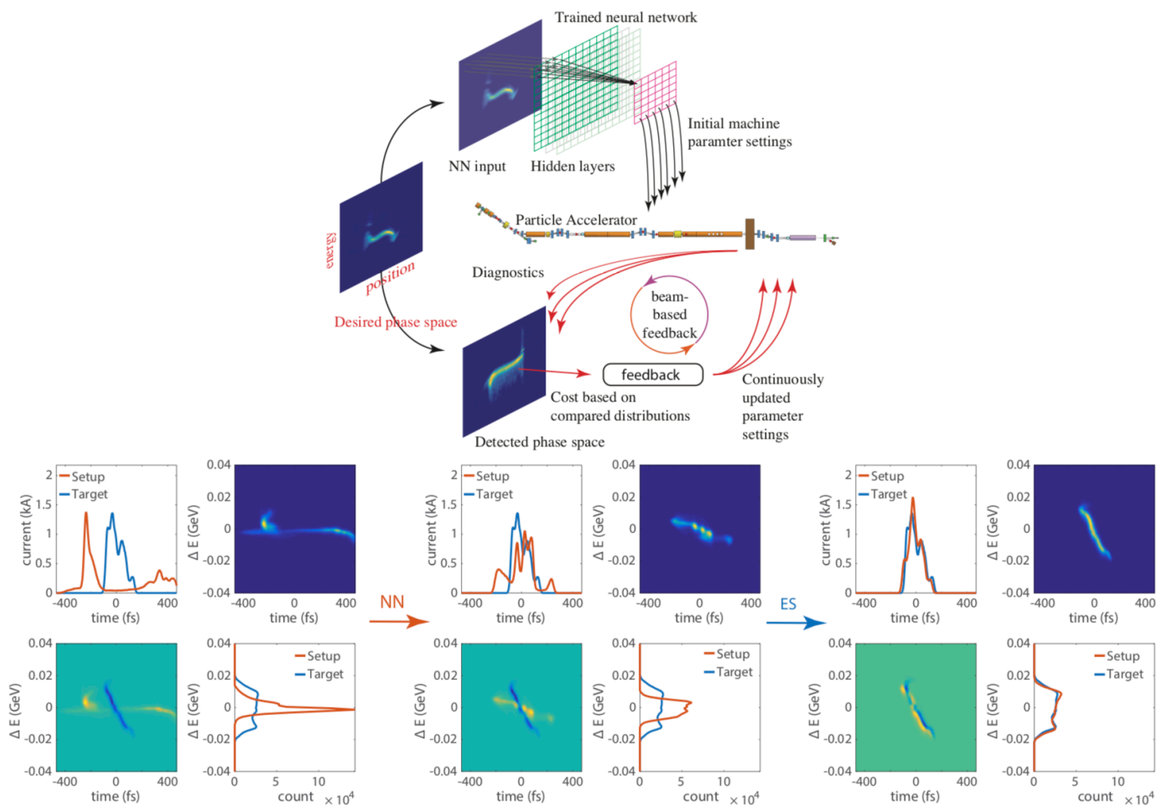
CMS L1 & High-
Level Triggers
50k cores, 1kHz

Growing Literature

https://inspirehep.net/literature?q=machine learning or deep learning

- Community-based up to date listing of references
  https://iml-wg.github.io/HEPML-LivingReview/ (please update and contribute)
- Impossible to cover and give credits to all aspects of AI-4-HEP in this talks .
- Below are **highlights on specific data handling related applications**.
- Think that something major is missing ? Let me know and I'll amend.

Producing the Data

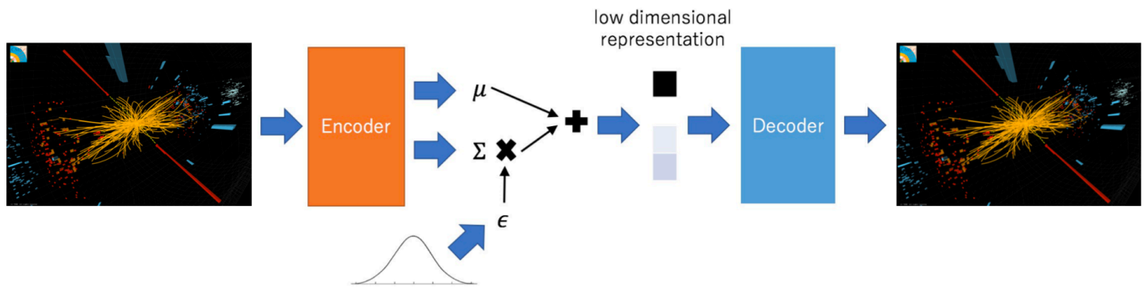


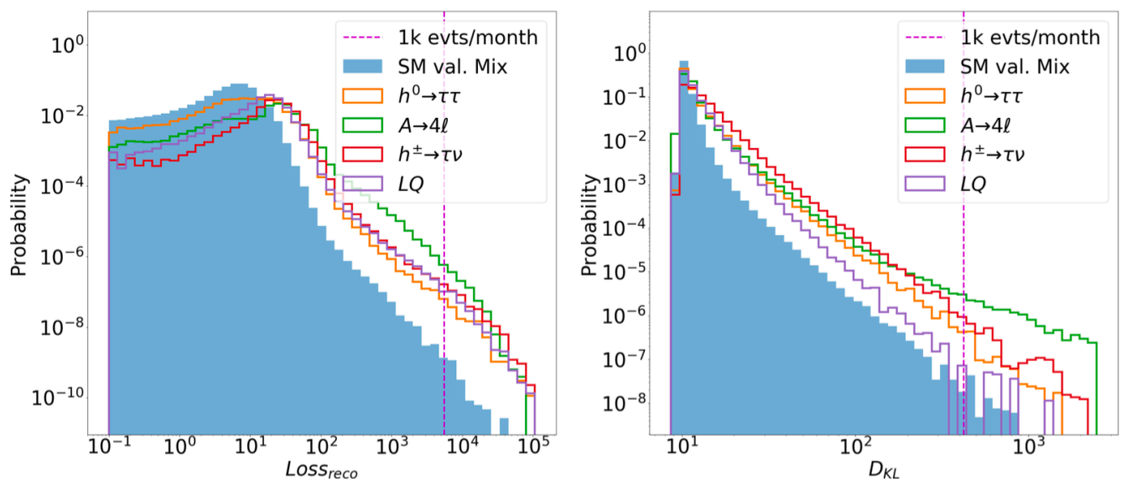
Advanced Control Methods for Particle Accelerators
https://arxiv.org/abs/2001.05461

- Machine learning can be used to tune devices, control beams, perform analysis on accelerator parameters, etc.
- Already successfully deployed on accelerator facilities.
- More promising R&D to increase beam time.

Opportunities in Machine Learning for Particle Accelerators https://arxiv.org/abs/1811.03172
Machine learning for design optimization of storage ring nonlinear dynamics https://arxiv.org/abs/1910.14220
Advanced Control Methods for Particle Accelerators (ACM4PA) 2019 Workshop Report https://arxiv.org/abs/2001.05461
Machine learning for beam dynamics studies at the CERN Large Hadron Collider https://arxiv.org/abs/2009.08109
…

Acquiring Data

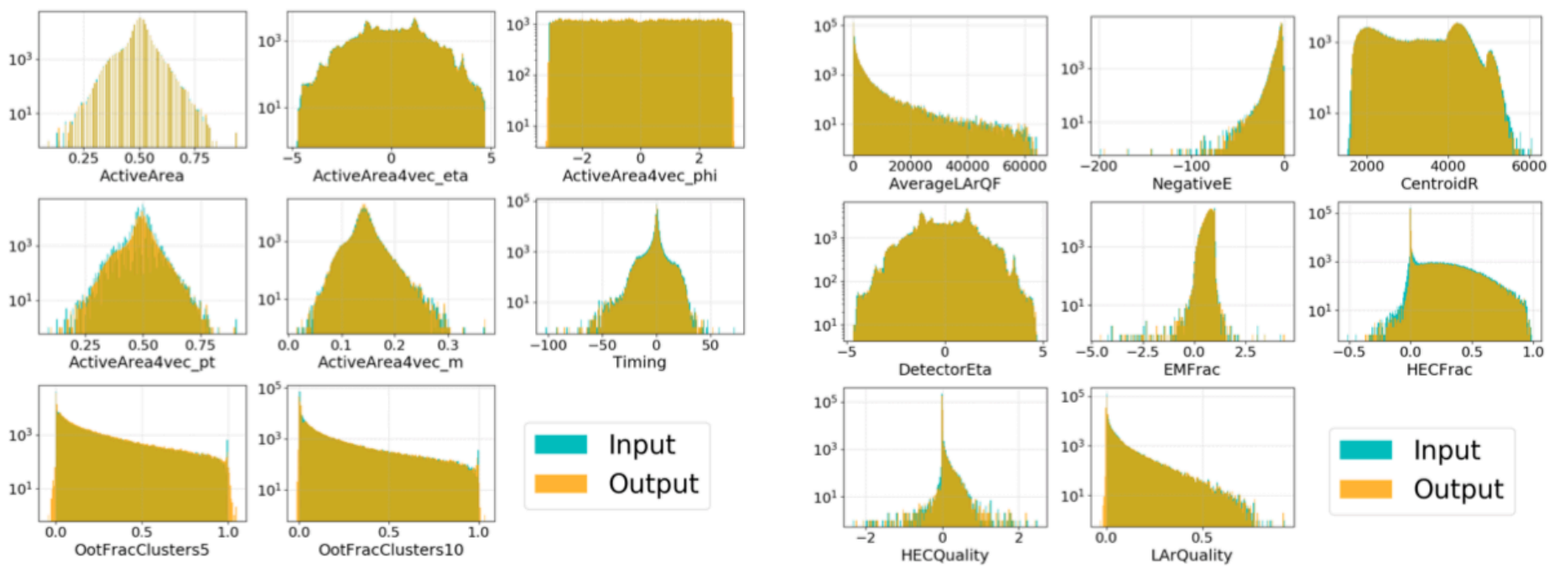




Use of variational auto-encoders directly on data to marginalize outlier events, for anomalous event hotline operation.
https://doi.org/10.1007/JHEP05(2019)036

- Machine learning since long deployed in the trigger for selected signatures.
- Further potential for background trigger rate reduction.
- Emerging opportunity for triggering on unknown signatures.
- More promising R&D and experiment adoption.

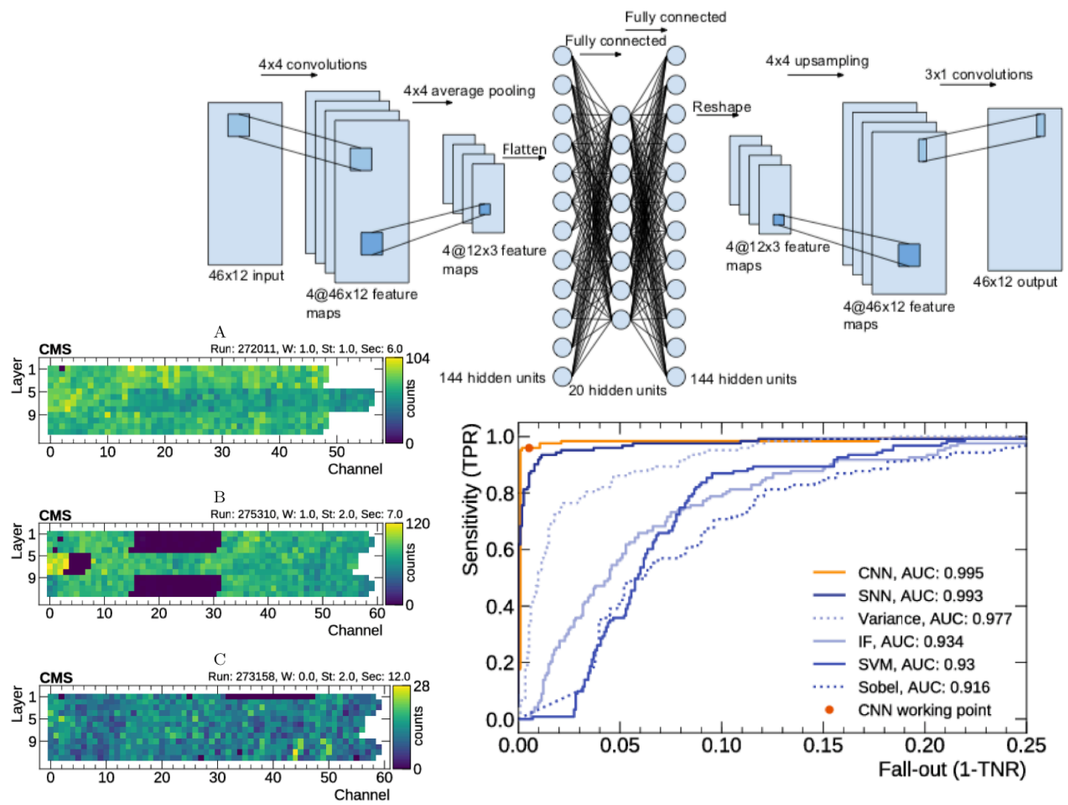
Compressing Data



Use of auto-encoder model
http://lup.lub.lu.se/student-papers/record/9004751

- Rich literature on data compression of image with neural network.
- Make use of abstract semantic space for image compression.
- Image compression can suffer some loss of resolution.
- Saving on disk/tape cost. Potential in scouting data analysis.
- R&D needed to reach the necessary level of fidelity.

Cleaning Data

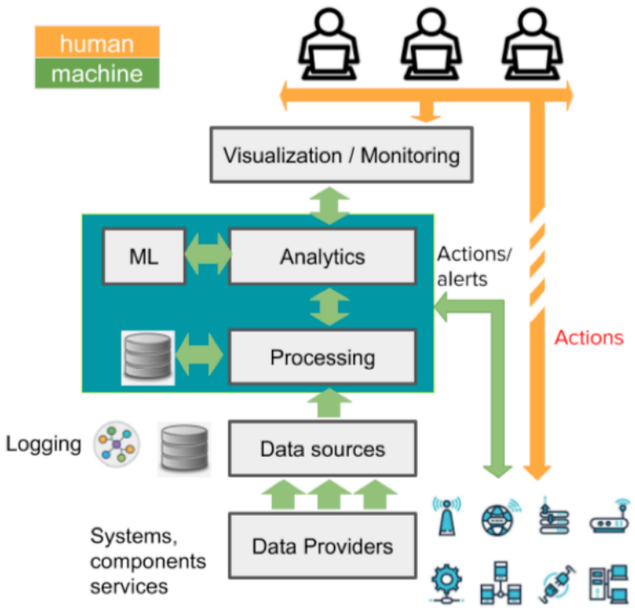


Use of auto-encoders to isolate anomalous monitoring histograms
https://doi.org/10.1007/s41781-018-0020-1

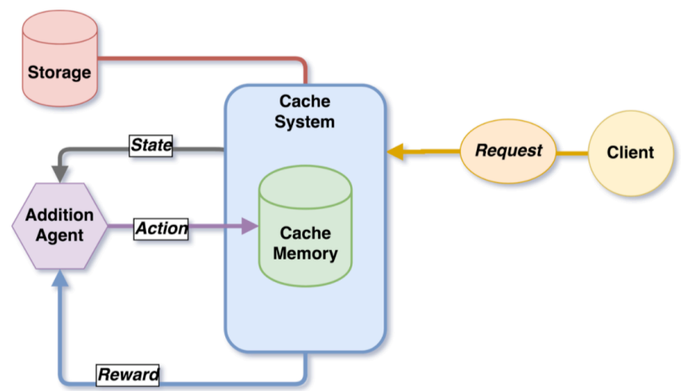
- Data quality is a person power intensive task, and crucial for swift delivery of Physics
- Machine learning can help with automation.
- Learning from operators, reducing workload.
- Continued R&D and experiment adoption.

Towards automation of data quality system for CERN CMS experiment https://doi.org/10.1088/1742-6596/898/9/092041
LHCb data quality monitoring http://dx.doi.org/10.1088/1742-6596/898/9/092027
Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider https://arxiv.org/abs/1808.00911
Anomaly detection using Deep Autoencoders for the assessment of the quality of the data acquired by the CMS experiment https://doi.org/10.1051/epjconf/201921406008
…

Managing Data



http://cds.cern.ch/record/2709338/
https://operational-intelligence.web.cern.ch

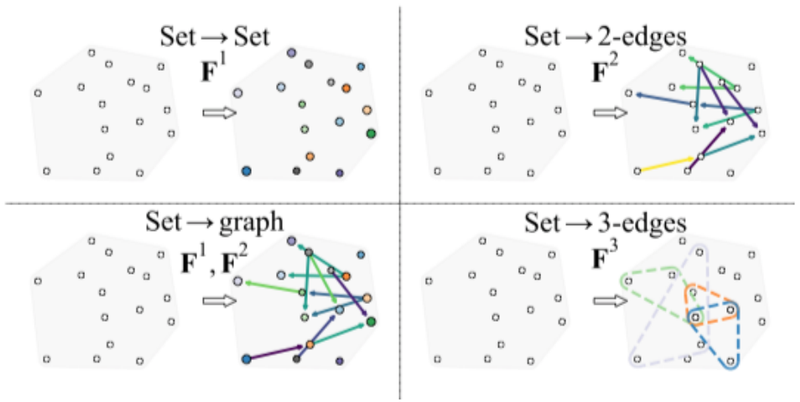


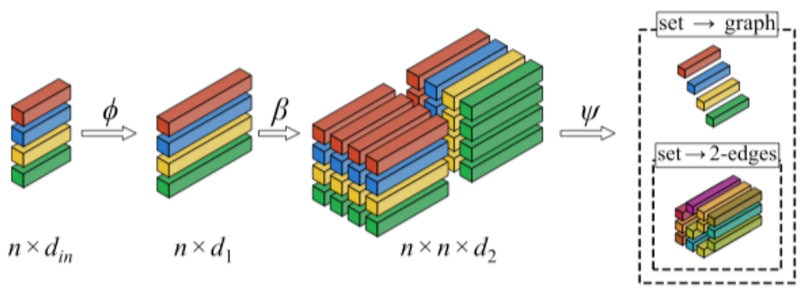
Cache Type	Throughput	Cost	Read on hit ratio	Band sat.	CPU Eff.
SCDL	**79.43%**	**50.68%**	21.22%	58.94%	58.75%
LFU	65.01%	104.73%	**33.29%**	**51.00%**	**60.92%**
Size Big	49.02%	111.73%	28.55%	54.40%	60.41%
LRU	47.15%	112.84%	27.64%	54.93%	59.90%
Size Small	46.71%	113.01%	27.39%	55.01%	59.73%

Caching suggestions using Reinforcement Learning
LOD 2020, in proceedings

- The LHC-grid is key to success of the LHC experiments.
- Complex ecosystem with dedicated operation teams.
- Person power demanding, and inefficient in some corner of the phase space.
- Potential for AI-aided operation.
- Lots of modeling and control challenges.
- R&D to increase operation efficiency.

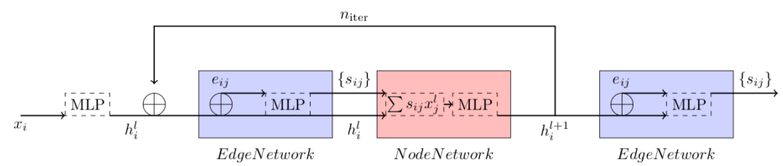
Reconstructing Data

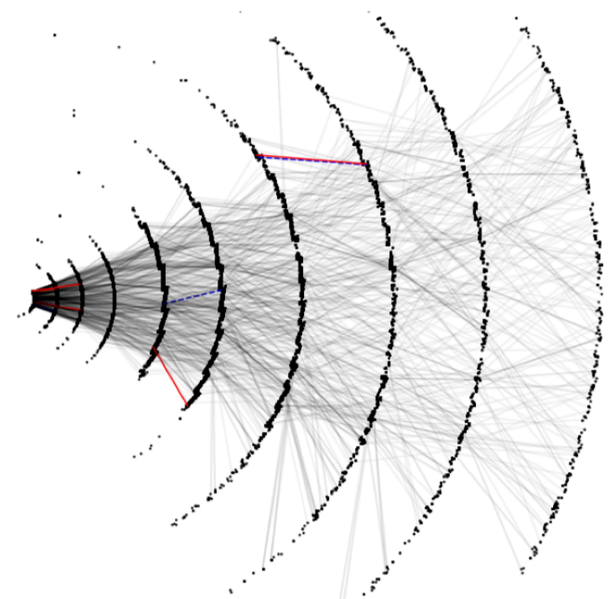




Learning graphs from sets, applied to vertexing
https://arxiv.org/abs/2002.08772

GNN applied to charged particle tracking
https://arxiv.org/abs/2007.00149

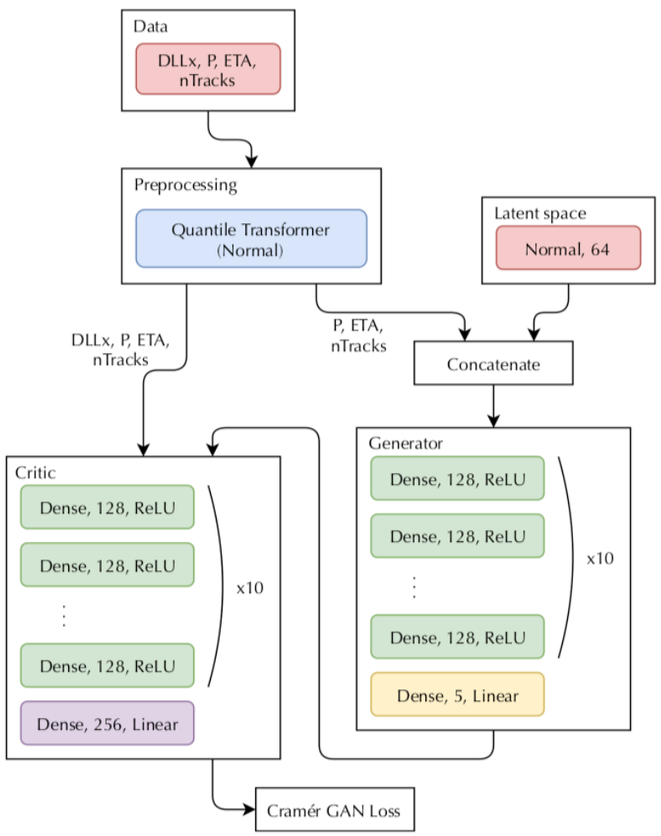


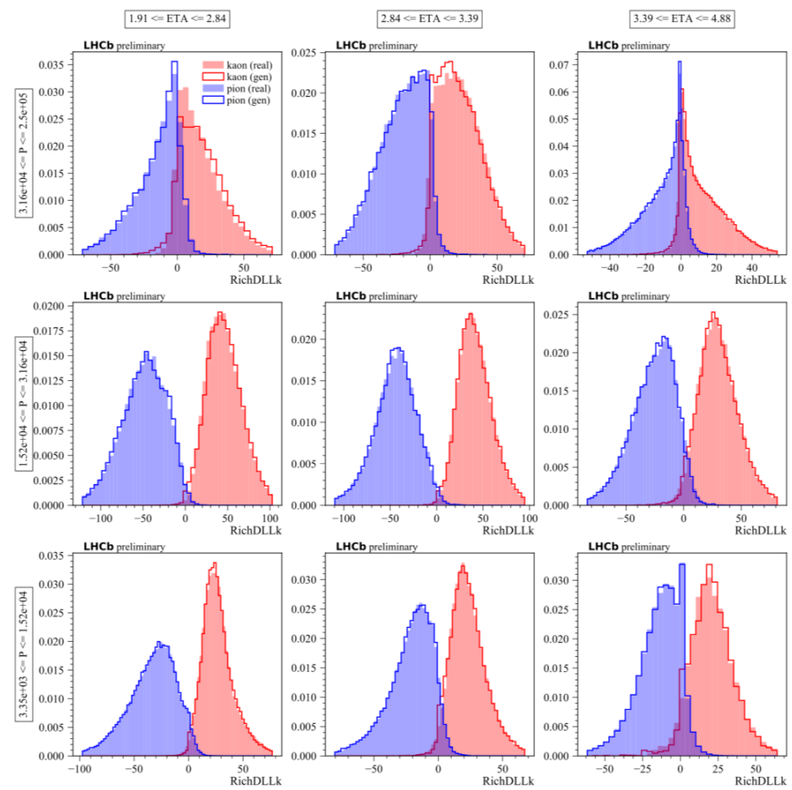


- Event reconstruction is pattern recognition to a large extend. Advanced machine learning techniques can help.
- Learn from the simulation, and/or data.
- Learn from existing "slow reconstruction" or simulation ground truth.
- Automatically adapt algorithm to new detector design.
- Image base methods evolving towards graph-based methods.
- Accelerating R&D to exploit full potential.

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

Simulating Data



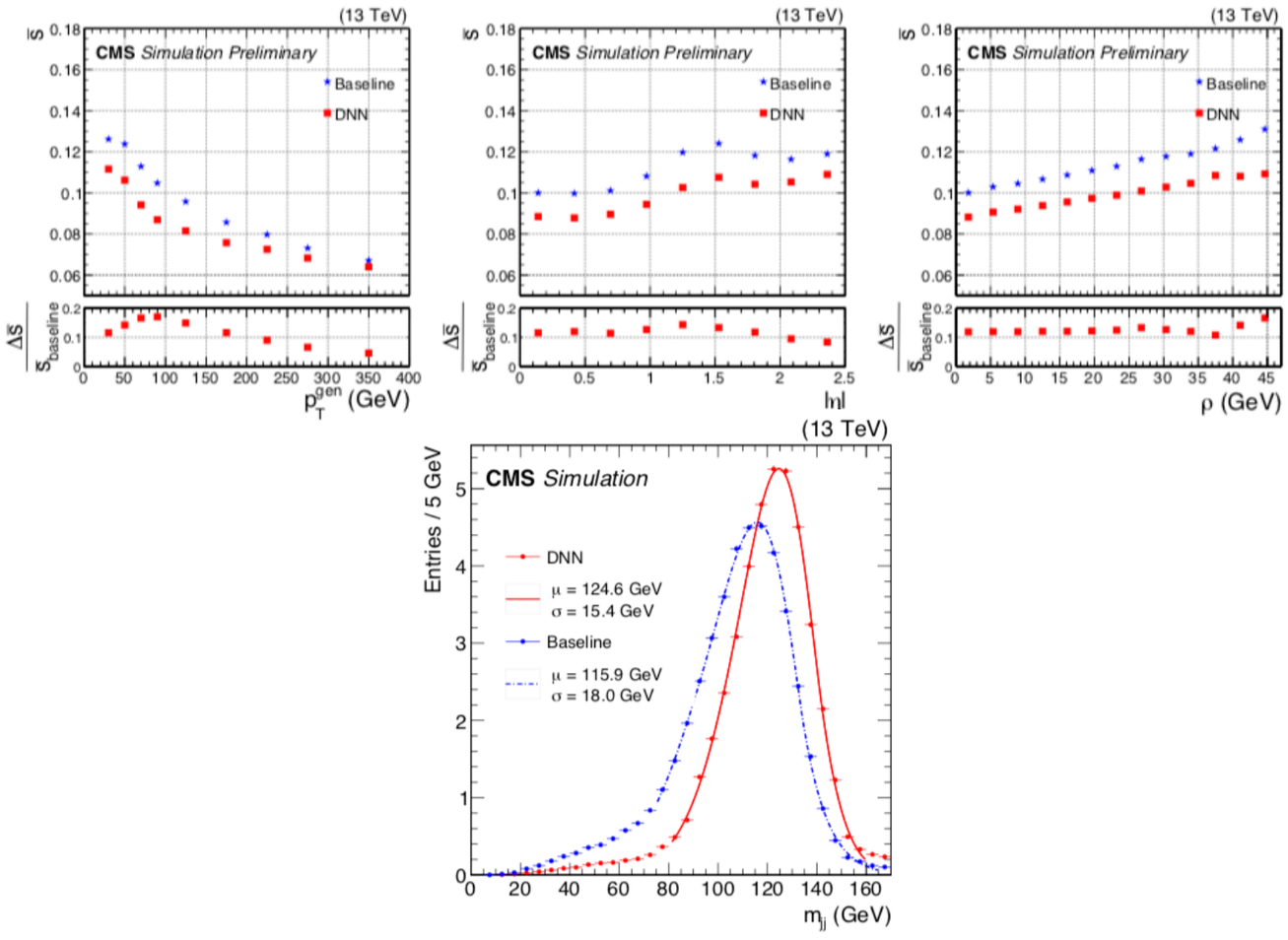


Generative Adversarial Networks for LHCb Fast Simulation
https://arxiv.org/abs/2003.09762

- Fully detailed simulation is computing intensive.
- Fast and approximate simulators already in operation.
- Applicable at many levels : sampling, generator, detector model, analysis variable, etc
- Generative models can provide multiple 1000x speed-up.
- Careful study of statistical power of learned models over training samples.
- Many R&D, experiment adoption starting.

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

Calibrating Data

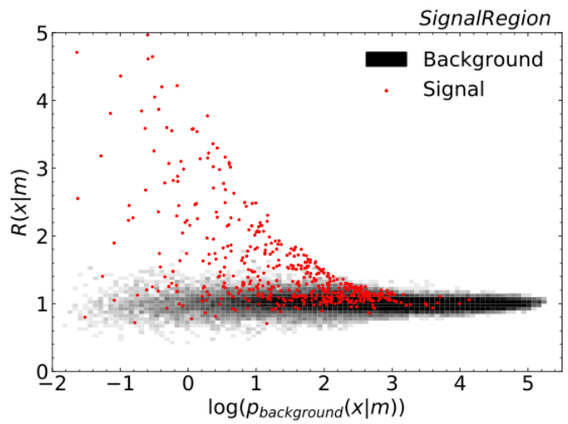


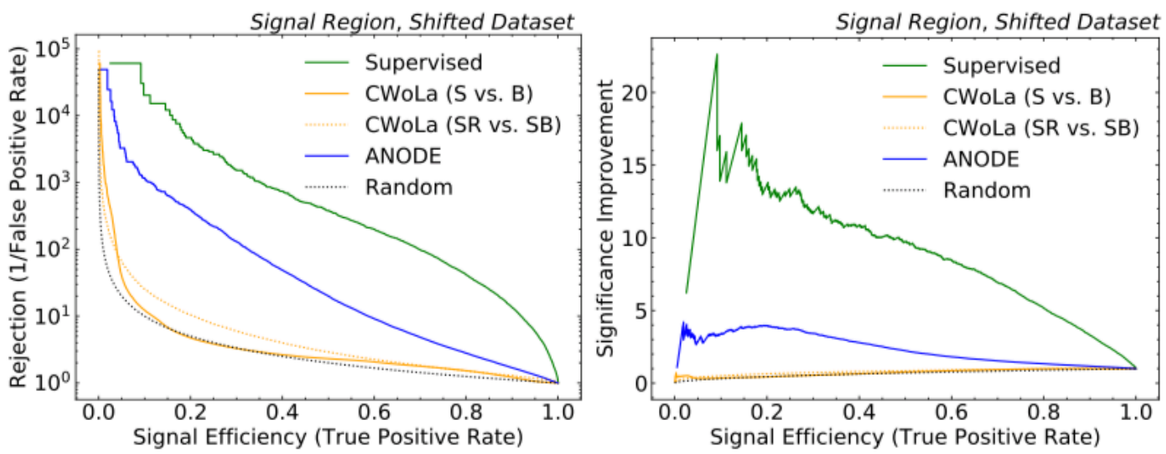
A deep neural network for simultaneous estimation of b jet energy and resolution
https://arxiv.org/abs/1912.06046

- Energy regression is the most obvious use case.
- Learning calibrating models from simulation and data.
- Parametrization of scale factors using neural networks.
- Reducing data/simulation dependency using domain adaptation.
- Continued R&D

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

Analyzing Data



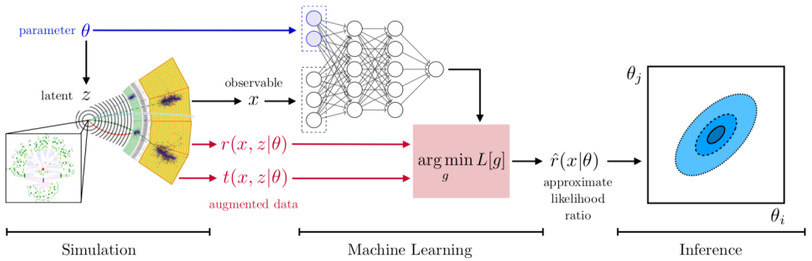


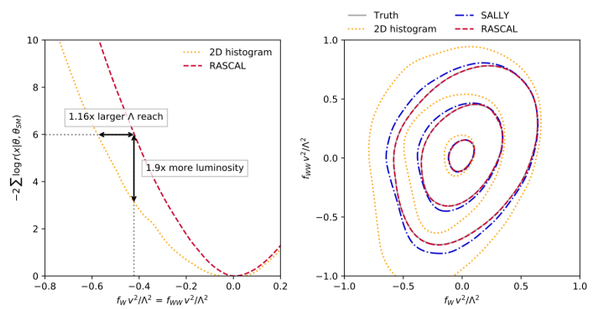
Use of masked autoregressive density estimator with normalizing flow as model-agnostic signal enhancement mechanism.
https://doi.org/10.1103/PhysRevD.101.075042

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

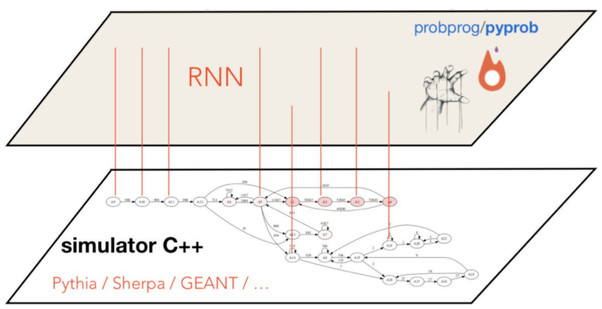
- Machine learning has long infiltrated analysis for signal/bkg classification.
- Increasing number of analysis with more complex DNN.
- Application to signal categorization, bkg modelling, kinematics reconstruction, decay product assignment, object identification, …
- Breadth of new model agnostic methods for NP searches.
- Continued R&D and experiment adoption initiated.

Theory Behind the Data

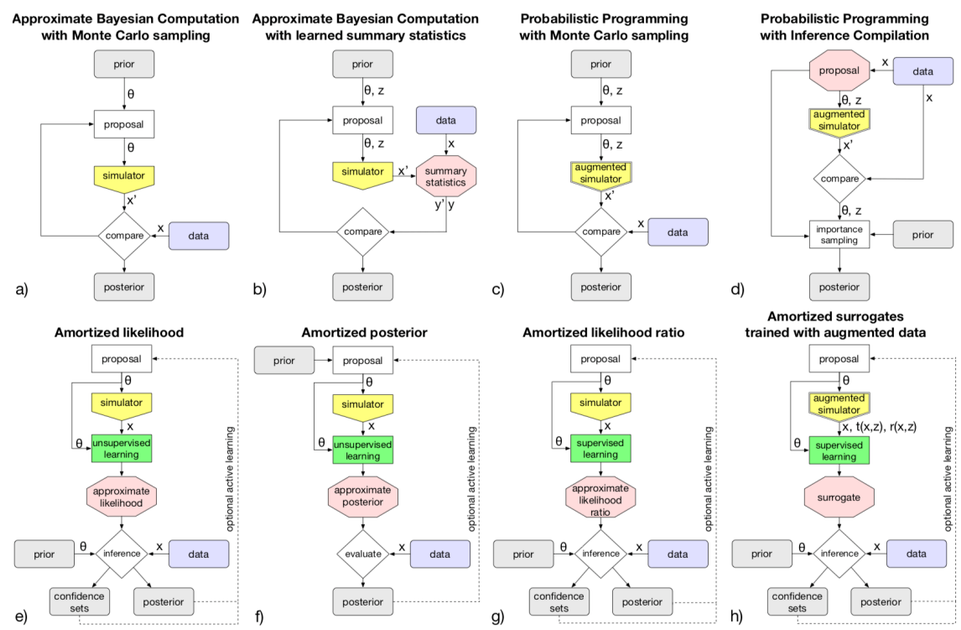




Constraining EFT with ML
https://arxiv.org/abs/1805.00013



https://github.com/probprog/pyprob



The frontiers of simulation-based inference
https://arxiv.org/abs/1911.01429

- Hypothesis testing is the core of HEP analysis.
- Intractable likelihood hinders solving the inverse problem.
- Going beyond the standard approach using machine learning and additional information from the simulator.
- More precise evaluation of the priors on theory's parameters.
- May involve probabilistic programming instrumentation of HEP simulator.
- R&D to bring this in the experiment.

Summary

- ➡Deep Learning (deep artificial neural networks) is dominating the field of machine learning nowadays
- ➡Deep learning can help in Science
- ➡Several deep learning solutions already deployed in HEP
- ➡Rapid growth of machine learning applications in HEP
  lots of details and applications left out in this talk
- ➡Turning proofs of concept into production (too) slowly
- ➡Exciting time ahead exploiting further the potential of AI

What Now

- ➡Machine learning education in HEP : complement, not supplement programming courses.
- ➡Recognition of dual DS-HEP profile in academia : keep the talents in the field.
- ➡Model training infrastructure: beyond institutional GPU-clusters.
- ➡Model deployment infrastructure : software/firmware integration, more than hardware planning.
- ➡Cross-experiment development platform : sharable dataset, sharable dataset, sharable dataset.

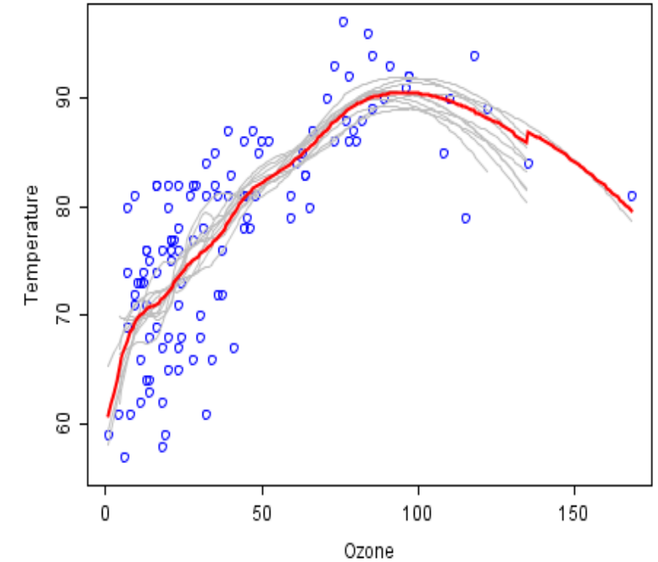
Extra material

Supervised Learning

- Given a dataset of samples, a subset of features is qualified as **target**, and the rest as **input**
- Find a **mapping from input to target**
- The mapping should **generalize to any extension** of the given dataset, provided it is generated from the same mechanism

$$dataset \equiv \{(x_i, y_i)\}_i$$
$$find \;\; function \;\; f \;\; s.t. \;\; f(x_i) = y_i$$

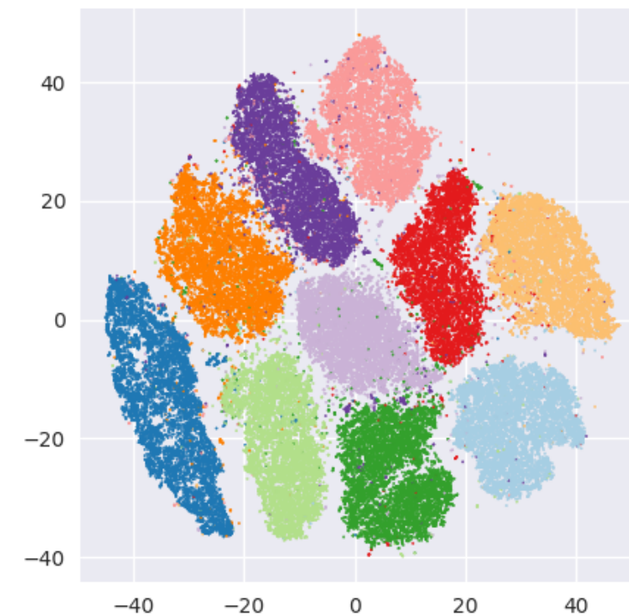
- Finite set of target values :
  - → **Classification**
- Target is a continuous variable :
  - → **Regression**



Unsupervised Learning

- Given a dataset of samples, but there is no subset of feature that one would like to predict
- Find mapping of the samples to a lower dimension manifold
- The mapping should generalize to any extension of the given dataset, provided it is generated from the same mechanism

$$dataset \equiv \{(x_i)\}_i$$
$$find \;\; f \;\; s.t. \;\; f(x_i) = p_i$$



- Manifold is a finite set
  - → **Clusterization**
- Manifold is a lower dimension manifold :
  - → **Dimensionality reduction, density estimator**

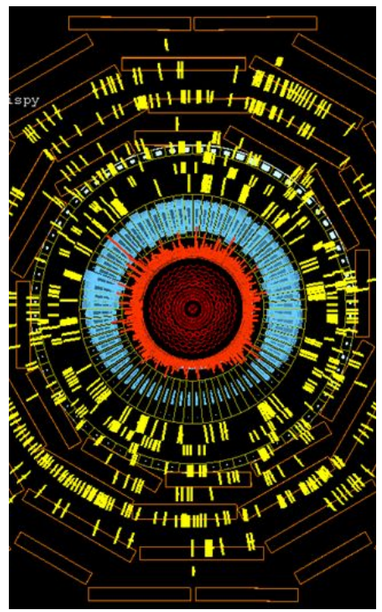
Reinforcement Learning

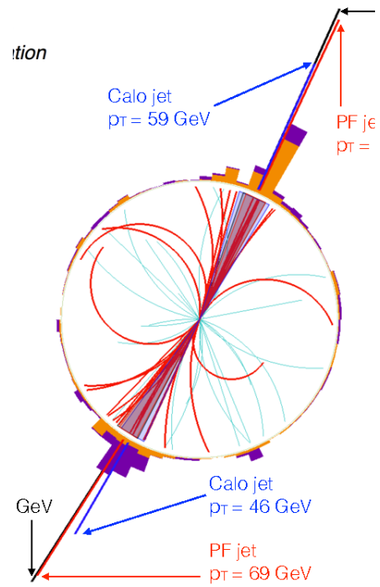
- Given an **environment** with multiple states, given a reward upon action being taken over a state
- Find an **action policy to drive** the environment toward maximum cumulative reward

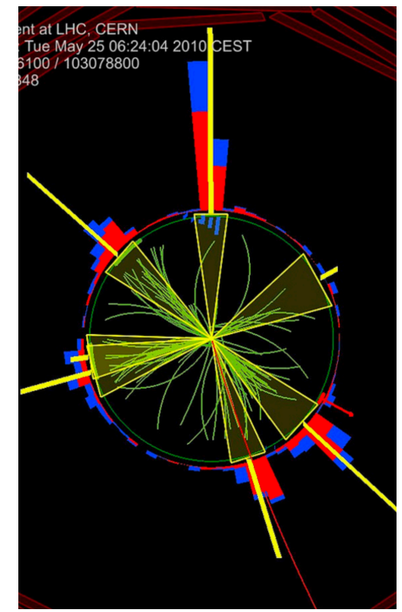
$$s_{t+1} = Env(s_t, a_t)$$
$$r_t = Rew(s_t, a_t)$$
$$\pi(a|s) = P(A_t = a|S_t = s)$$
$$find\ \pi\ s.t. \sum_t r_t\ is\ maximum$$

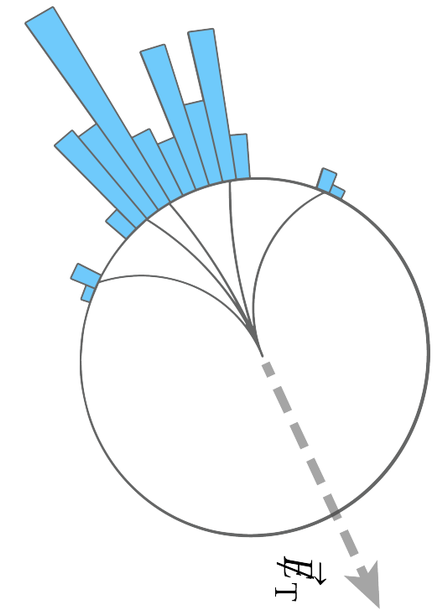
From RAW to High Level Features

Detector Data	Local reconstruction	Particle representation	Jet Clustering	High level features









Event Processing

Dimensionality reduction

Globalization of information

From digital signal, to local hits, to a sequence of particles, jets, and high-level features.
Complex and computing intensive task that could find a match in ML application.

Reconstruction

Mostly pattern recognition tasks with regressions and classifications
Development on multiple tasks:

- Local energy reconstruction
- Jet reconstruction
- Particle (flow) reconstruction
- Tracking
- Vertexing
- …

Composite reconstruction and end-2-end approaches.

Graph neural networks are emerging as overarching solutions for reconstruction tasks

Simulating Collisions

Madgraph, Pythia, Sherpa, ...	**Event Generator**: compute predictions of the standard models to several orders of expansion in coupling constants (LO, NLO, NNLO, ...) using proton density functions.
Pythia, ...	**Hadronization**: phenomenological model of the evolution of hadrons under the effect of QCD.
GEANT 4, GEANT V	**Material simulator**: transports all particles throughout meters of detector, using high resolution geometrical description of the materials.
Homegrown software	**Electronic emulator:** converts simulated energy deposits in sensitive material, into the expected electronic signal, including noise from the detector.

Non-differentiable sequence of complex simulators of the signal expected from the detectors.
Computing intensive task, exceeding budget for reconstruction.

Generative Models in HEP

Enormous potential gain in computing performance (x thousands)
Extensive R&D effort on-going

- Better PDF
- Faster phase space integration
- Matrix element surrogates
- Particle shower energy deposit simulation
- Analysis-level sample production
- Learning the detector transfer function
- Reconstructed-particle-level fast simulation
- …

Extension of such models

- Unfolding mechanism
- Anomaly detection
- Background subtraction
- …

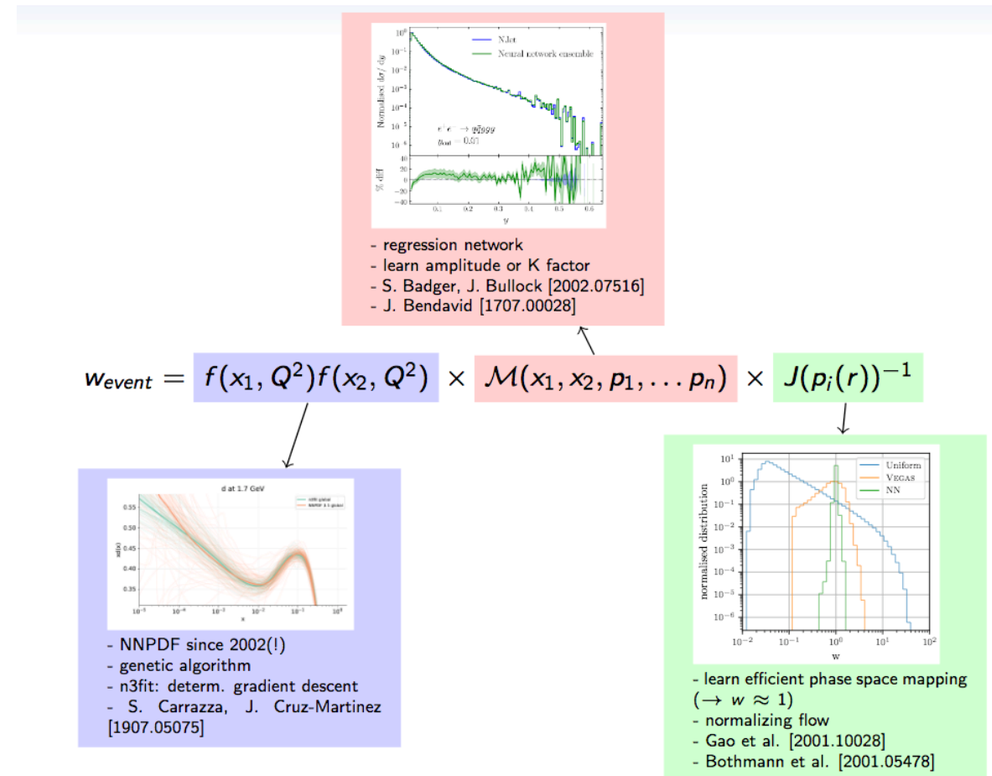


Diagram A. Butter

Generative adversarial network (GAN),Variational Auto-Encoder (VAE) and hybrid solutions taking the best of both methods

Taking Control of Apparatus

Multiple components of HEP data pipe-line can be operated with AI

- Accelerators
- Detectors
- DAQ/trigger
- Internet networking
- Data management
- Computing facilities
- …

Reinforcement learning for learning policies is challenging to put in practice (environment simulation / lots of data required, …).
Alternative approaches using supervised learning.

Potential gain in operation cost, and utilization efficiency.