

BNL High Energy Physics AI/ML vision

September 2021

Machine learning techniques have been successfully used in HEP and other sciences for over 30 years. The deep learning revolution since a decade ago and consequent rapid growth in AI/ML techniques, accompanying software tools and leveraged hardware is creating tremendous potential for enhancement in AI/ML scientific computing tools with multiple benefits to HEP.

Our vision for AI/ML is two-fold. We seek to exploit the development of established and emerging AI/ML techniques to solve experiment or theory driven computing challenges, and we seek to develop AI/ML techniques assuring that the application of AI/ML to our problems is sound, with quantifiable systematic uncertainties. We see AI/ML as a tool and approach to apply after exhausting well understood "conventional" ways to extract information from data. With this strategy we use accrued domain knowledge to wisely guide the application of AI/ML, to produce results that can be understood and trusted, together with their error properties.

AI/ML is being widely investigated and used today in BNL HEP programs, including neutrino physics, ATLAS, astroparticle physics and HEP theory. While a rapidly evolving area, AI/ML has been demonstrated and accepted as an important tool capable of improving our physics results. Today's AI/ML activities and near term plans in HEP are a mix of exploratory R&D, concrete applications, tool development, and planning/developing towards future applications. All research programs have identified current AI/ML activities in separate FWPs. Current AI/ML applications and research areas include signal reconstruction and event topology for neutrino events; event reconstruction, physics analysis and trigger/DAQ applications in ATLAS; ML studies of Higgs production including next to leading order QCD for anomalous Higgs couplings; and the development of highly scalable ML services to provide processing that empowers scientific creativity in developing sophisticated AI/ML applications.

An example of work delivering in both the near and long term is in ATLAS simulation. At the HL-LHC ATLAS aims to use fast simulation for 90% of its simulation statistics, saving time relative to full Geant4 but making fast simulation a substantial processing consumer in itself. The fast simulation processing chain is projected to use 21% of ATLAS processing in 2030. Using ML to improve both the speed and accuracy of the fast simulation is making good progress. BNL-led work on speeding up the FastCaloGAN calorimeter simulation, the first ML based simulation code to enter ATLAS production, has lately improved training speed by 2-3x and evaluation speed by 5x. ML

based simulation can leverage large heterogeneous HPCs, which are increasingly designed and optimized for good AI/ML performance. With one FastCaloGAN training cycle taking ~1 GPU-month of processing time, ATLAS is already presenting significant ML processing needs that can benefit from large scale resources.

In an example of applying AI/ML to the understanding of simulation systematics, a new R&D program at BNL makes use of generative adversarial networks (GANs) to train a model that translates simulated HEP data to real data. A given simulated event will then differ from its translated version in ways that can reveal real inaccuracies, artifacts, resolutions and biases in the simulation. Furthermore, with the difference between simulated events and their translated form representing the modelling deficiency with respect to actual data, the simulated events and their translated counterparts can be propagated through the full offline software chain with the difference between final outputs representing the overall detector simulation systematic error. The data volumes and neural networks involved are large, requiring large scale processing for training and inference that can benefit from HPCs and the BNL-developed services to use them easily.

BNL HEP AI/ML activities are strengthened and complemented by activities elsewhere in the Laboratory, including Nuclear Physics; an example is a new R&D collaboration between ATLAS and Electron Ion Collider (EIC) physicists to apply AI/ML intelligence at the detector. With LHC data rates at the PB/s scale, annual processed data volumes already exceeding an exabyte, and storage the largest cost component of LHC computing (and inevitably large for future facilities as well), intelligent filtering at the experiment's DAQ system can potentially yield large economies in downstream costs without degrading the physics. BNL is applying its TDAQ expertise to develop this approach, using modern deep learning techniques on powerful commodity FPGAs to achieve ~100x reduction in data storage requirements.

As we look to our AI/ML vision, our focus in the near term (1-2 years) is on building up community, scientific applications and the tools to empower them. AI/ML coordination within HEP and at Laboratory level facilitates communication, collaboration and 'training up' the staff.

Example activities in this period: AI/ML is an established and growing contributor to faster and more accurate ATLAS simulation, with BNL-developed distributed AI/ML services facilitating the use of large scale heterogeneous HPCs for simulation production. The AI/ML based study of simulation systematics will be applied to the first target application, ProtoDUNE LAr TPC simulation.

Our mid term (3-5 year) focus will be on the maturation of current and emerging R&D, and Laboratory infrastructure integration. R&D topics being pursued on this timescale via Laboratory-directed R&D (LDRD), DOE's SciDAC program and other mechanisms include the mentioned simulation systematics study (LDRD) and AI/ML intelligence at the detector (LDRD), real-time particle tracking with deep learning on field programmable gate arrays (FPGAs) (LDRD), image classification in the HEP cosmic frontier (SciDAC), an Intelligent Data Delivery Service supporting large scale AI/ML workflows (US ATLAS HL-LHC computing R&D), and AI/ML based operational intelligence in workflow management (ATLAS computing R&D). Most of these studies benefit from strong engagement with various organizations at BNL and beyond.

Example activities in this period: The Laboratory will deploy a common computing infrastructure supporting AI/ML with which we plan to integrate HEP activities, including our large scale AI/ML workflow services serving applications running on BNL infrastructure, DOE HPCs and other facilities. Applications will include ATLAS simulation, and simulation systematics studies extending beyond ProtoDUNE to other experiments (e.g. sPHENIX). The DAQ collaboration with EIC aims to deliver an AI/ML based tool chain implemented in firmware and software to optimize datataking in response to detector conditions.

Our longer term (5-10 year) vision focuses on consolidating BNL HEP as a leader in using AI/ML to increase the quality and productivity of scientific output across HEP programs, and in making BNL HEP a leader in the curation and use of high energy datasets for AI/ML and other studies. BNL will be a steward of crucial data-intensive experimental datasets (ATLAS including HL-LHC, Belle II, RHIC, EIC) that will be a resource for scientists employing AI/ML and other techniques for ongoing scientific discovery, facilitated by processing services that build on our distributed AI/ML services work, and data & analysis preservation work for RHIC. The EIC will be a rich source for collaboration on AI/ML applications, including streaming DAQ, with the R&D work getting underway today.