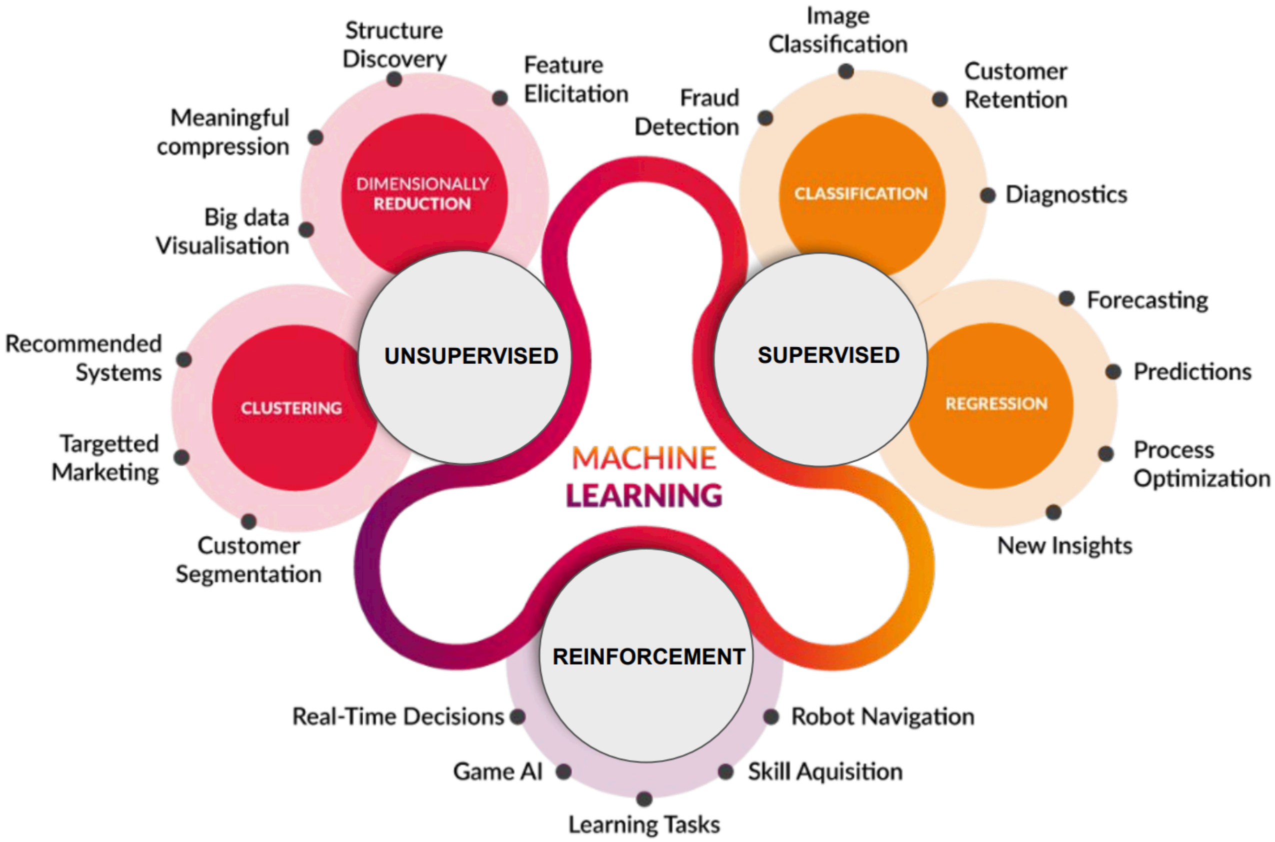
AI at Fermilab

Nhan Tran
January 25, 2024


Structure Discovery
Feature Elicitation
Meaningful compression
DIMENSIONALLY REDUCTION
Big data Visualisation
Image Classification
Customer Retention
Fraud Detection
CLASSIFICATION
Diagnostics
Forecasting
Predictions
REGRESSION
Process Optimization
New Insights
Recommended Systems
CLUSTERING
Targetted Marketing
Customer Segmentation
UNSUPERVISED
SUPERVISED
MACHINE LEARNING
REINFORCEMENT
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks

DOE HEP builds and operates among the most difficult and biggest projects with the most complex devices in science -- accelerators and detectors. Our priority is using AI for real-time controls, operations, and data processing to **accelerate scientific discovery at unprecedented data scales while creating enabling technology for society**

DOE HEP builds and operates among the most difficult and biggest projects with the most complex devices in science -- accelerators and detectors. Our priority is using AI for real-time controls, operations, and data processing to **accelerate scientific discovery at unprecedented data scales while creating enabling technology for society**

Algorithms for HEP science

Robust & generalizable learning; Fast & efficient algorithms; Adaptive & automated learning

Intelligent sensing and real-time processing

High performance and throughput compute

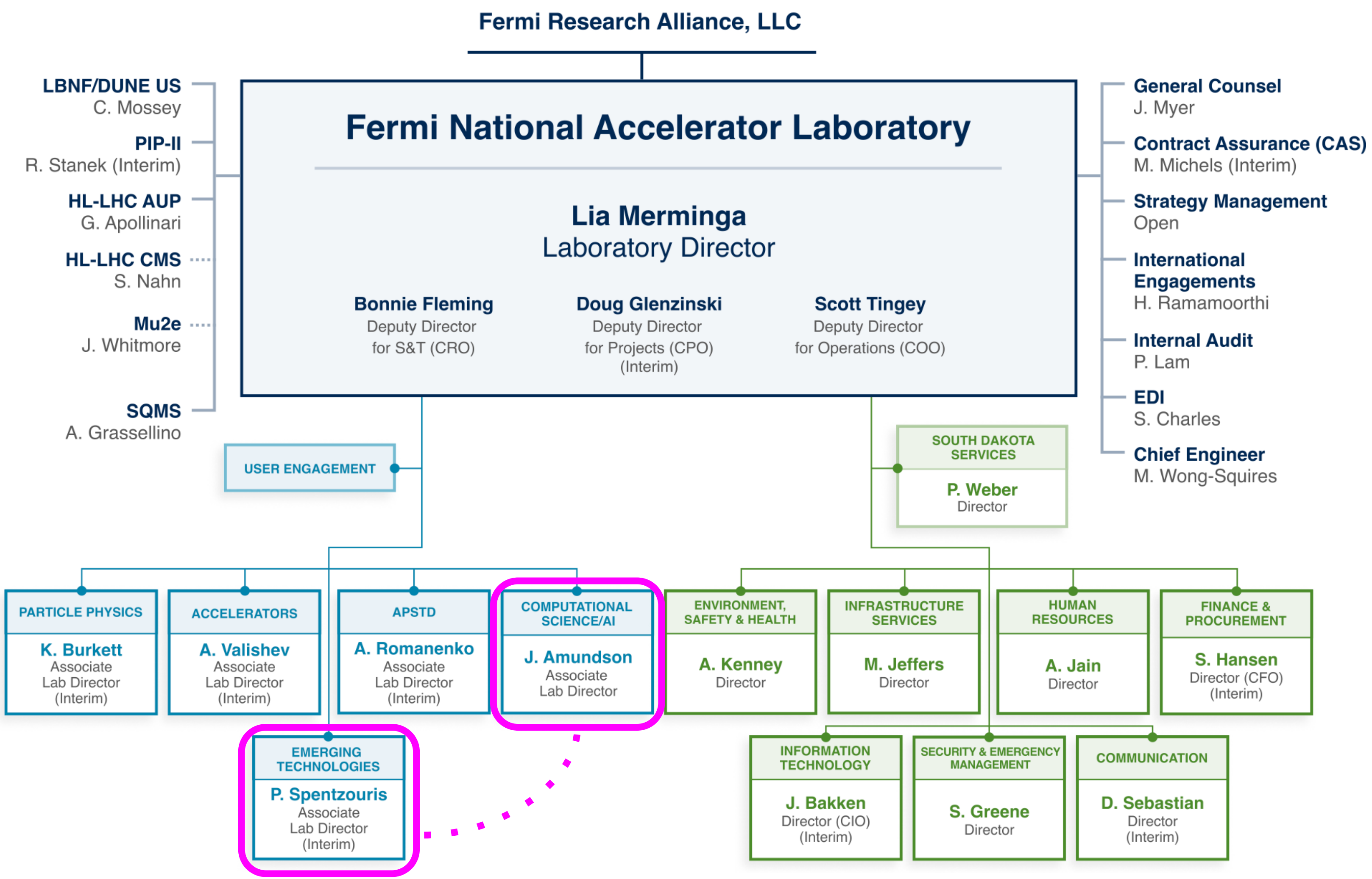
Operations, controls, analysis

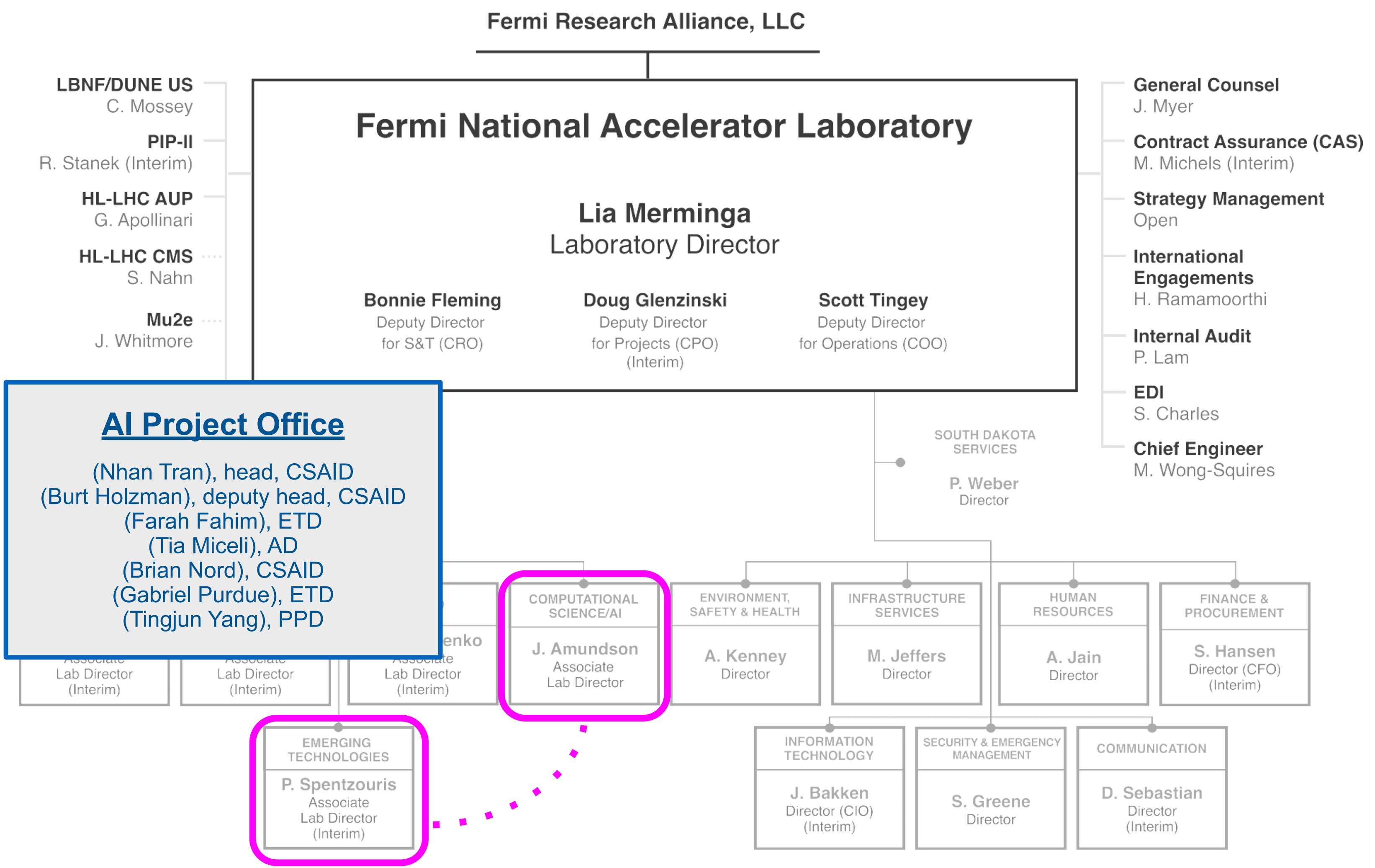
AI for HEP

Drivers to accelerate discovery

- **Deeper insights & better performance**
  Maximize science by getting the most out of machines and experiments; reduce systematics and understand anomalies
- **Accelerate time-to-physics**
  Enable powerful/robust ML at each stage of data processing; mitigate computing and data analysis challenges; automate scientific method and discovery
- **Improve operational efficiency**
  Optimize experimental “control” via triggers, data monitoring; recover lost data and physics

AI organization at Fermilab


Fermi Research Alliance, LLC
Fermi National Accelerator Laboratory
Lia Merminga
Laboratory Director
Bonnie Fleming
Deputy Director
for S&T (CRO)
Doug Glenzinski
Deputy Director
for Projects (CPO)
(Interim)
Scott Tingey
Deputy Director
for Operations (COO)
LBNF/DUNE US
C. Mossey
PIP-II
R. Stanek (Interim)
HL-LHC AUP
G. Apollinari
HL-LHC CMS
S. Nahn
Mu2e
J. Whitmore
SQMS
A. Grassellino
General Counsel
J. Myer
Contract Assurance (CAS)
M. Michels (Interim)
Strategy Management
Open
International Engagements
H. Ramamoorthi
Internal Audit
P. Lam
EDI
S. Charles
Chief Engineer
M. Wong-Squires
USER ENGAGEMENT
SOUTH DAKOTA SERVICES
P. Weber
Director
PARTICLE PHYSICS
K. Burkett
Associate Lab Director (Interim)
ACCELERATORS
A. Valishev
Associate Lab Director (Interim)
APSTD
A. Romanenko
Associate Lab Director (Interim)
COMPUTATIONAL SCIENCE/AI
J. Amundson
Associate Lab Director
EMERGING TECHNOLOGIES
P. Spentzouris
Associate Lab Director (Interim)
ENVIRONMENT, SAFETY & HEALTH
A. Kenney
Director
INFRASTRUCTURE SERVICES
M. Jeffers
Director
HUMAN RESOURCES
A. Jain
Director
FINANCE & PROCUREMENT
S. Hansen
Director (CFO) (Interim)
INFORMATION TECHNOLOGY
J. Bakken
Director (CIO) (Interim)
SECURITY & EMERGENCY MANAGEMENT
S. Greene
Director
COMMUNICATION
D. Sebastian
Director (Interim)

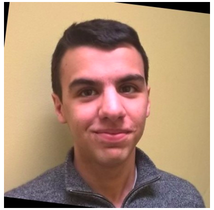
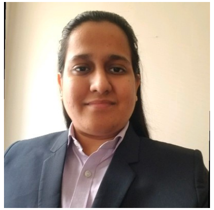


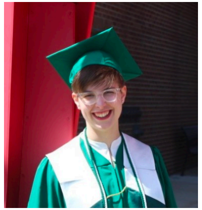
Fermi Research Alliance, LLC
Fermi National Accelerator Laboratory
Lia Merminga
Laboratory Director
Bonnie Fleming
Deputy Director
for S&T (CRO)
Doug Glenzinski
Deputy Director
for Projects (CPO)
(Interim)
Scott Tingey
Deputy Director
for Operations (COO)
LBNF/DUNE US
C. Mossey
PIP-II
R. Stanek (Interim)
HL-LHC AUP
G. Apollinari
HL-LHC CMS
S. Nahn
Mu2e
J. Whitmore
General Counsel
J. Myer
Contract Assurance (CAS)
M. Michels (Interim)
Strategy Management
Open
International Engagements
H. Ramamoorthi
Internal Audit
P. Lam
EDI
S. Charles
Chief Engineer
M. Wong-Squires
SOUTH DAKOTA SERVICES
P. Weber
Director
AI Project Office
(Nhan Tran), head, CSAID
(Burt Holzman), deputy head, CSAID
(Farah Fahim), ETD
(Tia Miceli), AD
(Brian Nord), CSAID
(Gabriel Purdue), ETD
(Tingjun Yang), PPD
Lab Director
(Interim)
Lab Director
(Interim)
enko
Lab Director
(Interim)
COMPUTATIONAL SCIENCE/AI
J. Amundson
Associate
Lab Director
ENVIRONMENT, SAFETY & HEALTH
A. Kenney
Director
INFRASTRUCTURE SERVICES
M. Jeffers
Director
HUMAN RESOURCES
A. Jain
Director
FINANCE & PROCUREMENT
S. Hansen
Director (CFO)
(Interim)
EMERGING TECHNOLOGIES
P. Spentzouris
Associate
Lab Director
(Interim)
INFORMATION TECHNOLOGY
J. Bakken
Director (CIO)
(Interim)
SECURITY & EMERGENCY MANAGEMENT
S. Greene
Director
COMMUNICATION
D. Sebastian
Director
(Interim)

Workforce development

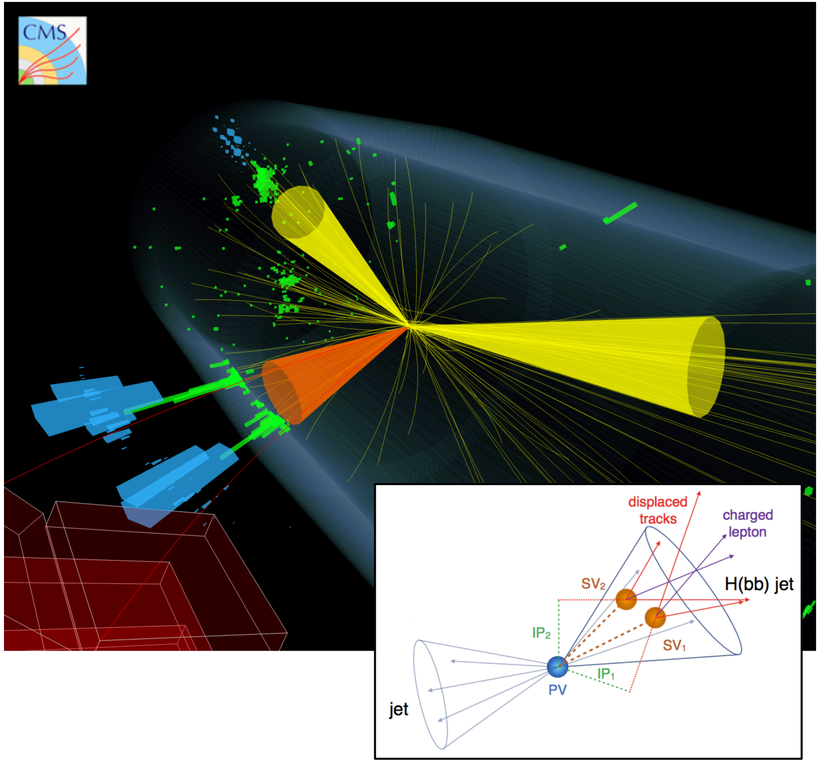
- New job type developed for AI research: **AI associate program**
  - New job family for advancement at Fermilab
- Modeled after industry 1-year internships
- Provides scientific AI research opportunities
  - Primarily Bachelors/MS with background in computer science & AI
- Concept emulated in other areas - e.g. engineering, quantum

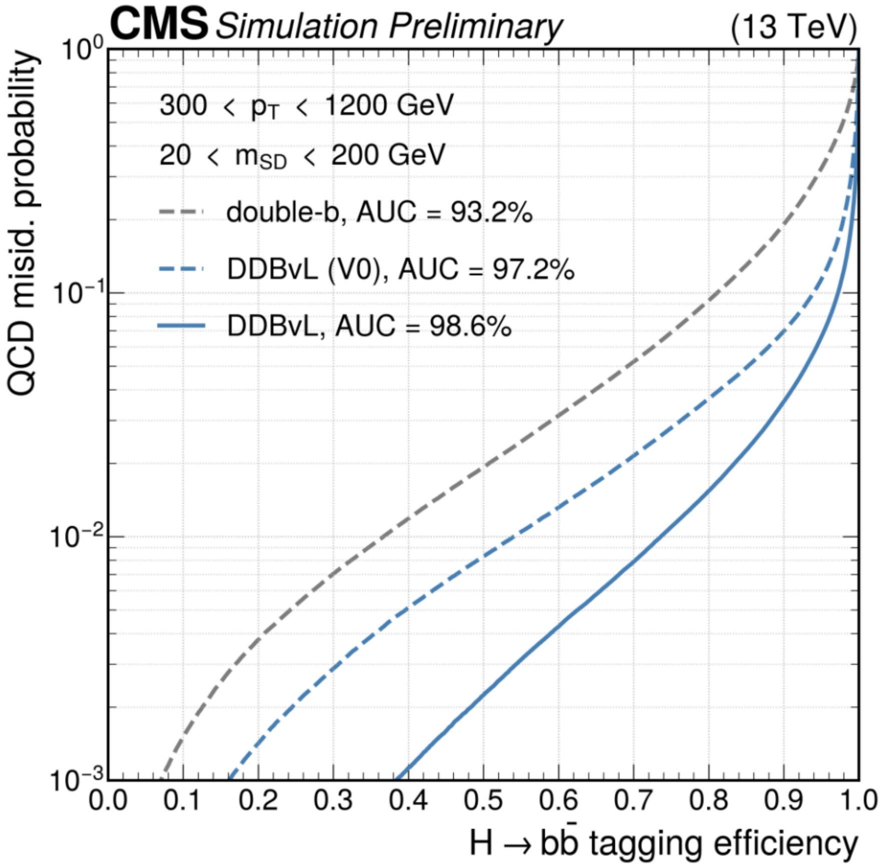


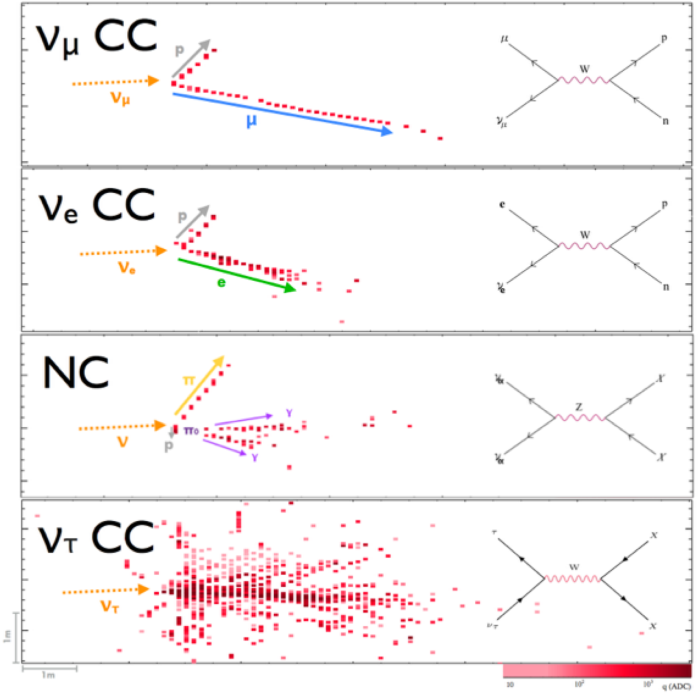



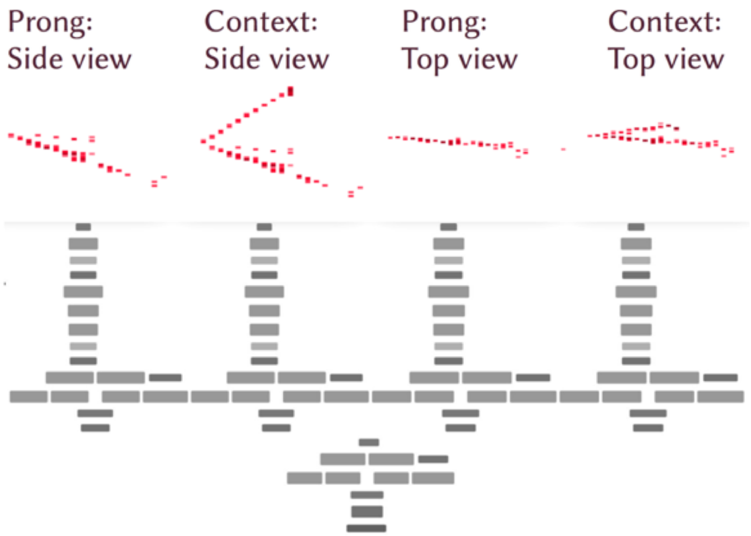
Selected AI highlights


CMS
displaced tracks
charged lepton
H(bb) jet
SV2
SV1
IP2
IP1
PV
jet


CMS Simulation Preliminary
(13 TeV)
300 < pT < 1200 GeV
20 < mSD < 200 GeV
double-b, AUC = 93.2%
DDBvL (V0), AUC = 97.2%
DDBvL, AUC = 98.6%
QCD misid. probability
H → bb̄ tagging efficiency
10^0
10^-1
10^-2
10^-3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

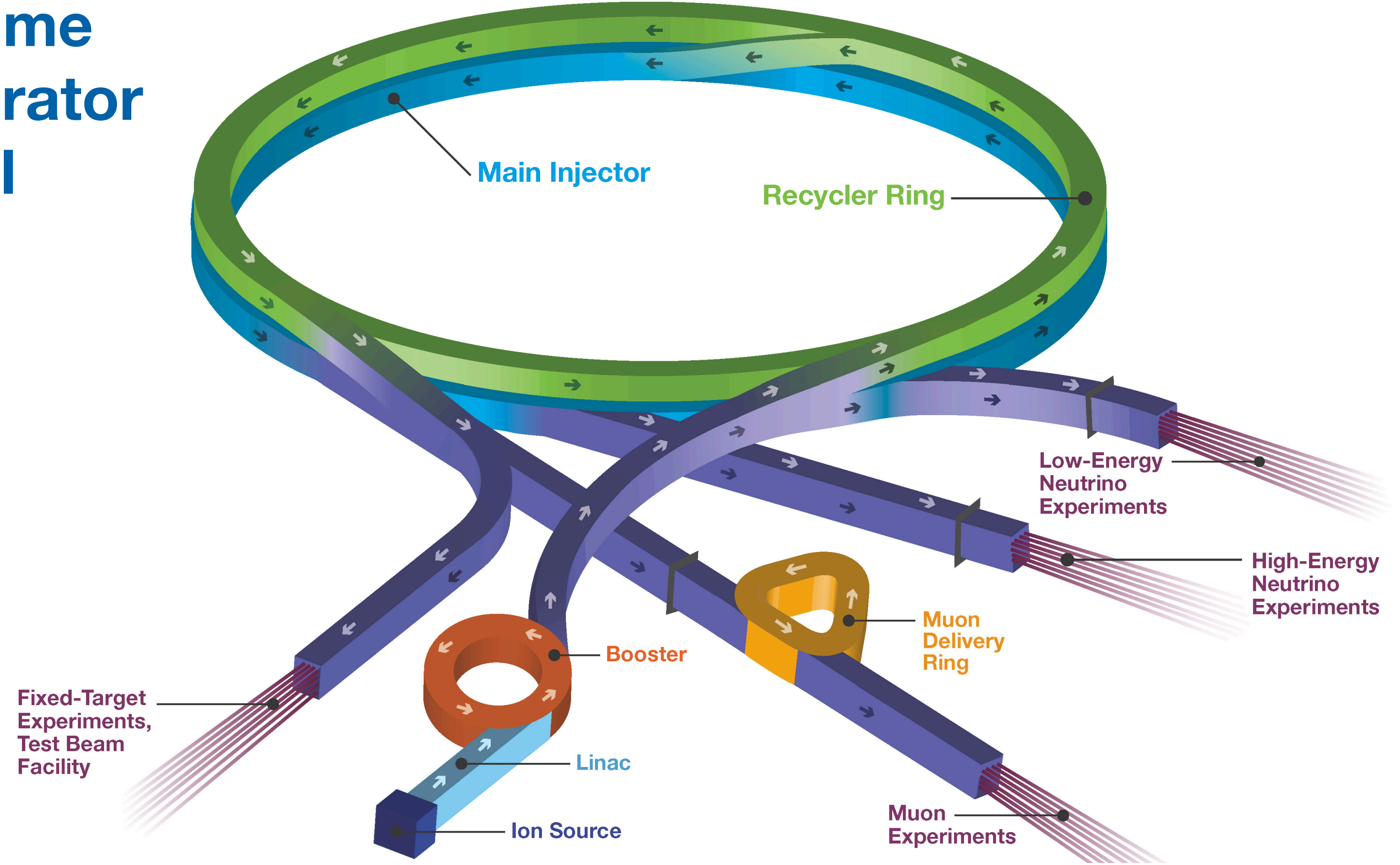
arxiv:1703.03328





Performance improvement equivalent to **4.2 kilotons of additional detector mass** with traditional particle identification algorithms.

Real-time accelerator control

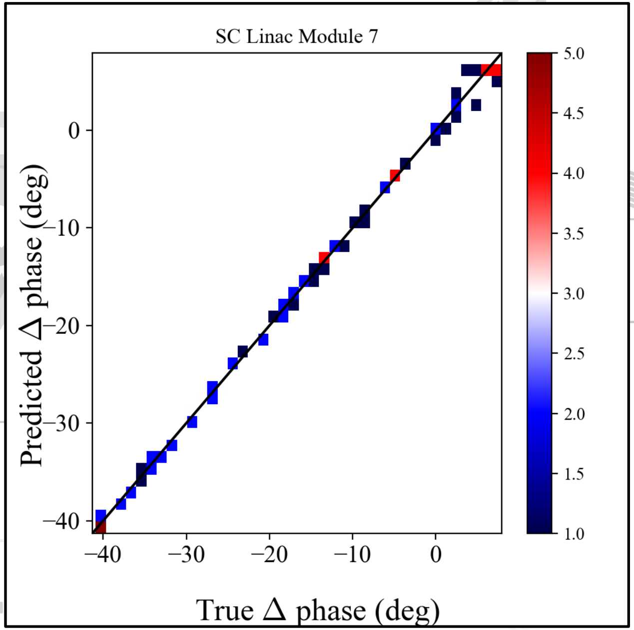


Real-time accelerator control

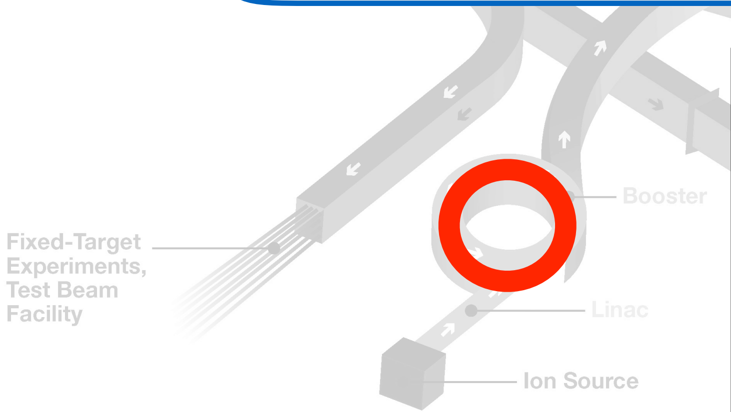
Linac RF optimization

Predict RF parameters to keep beam energy constant and minimize emmitance

Proof-of-concept with single cavity phase regulation; multi-cavity promising



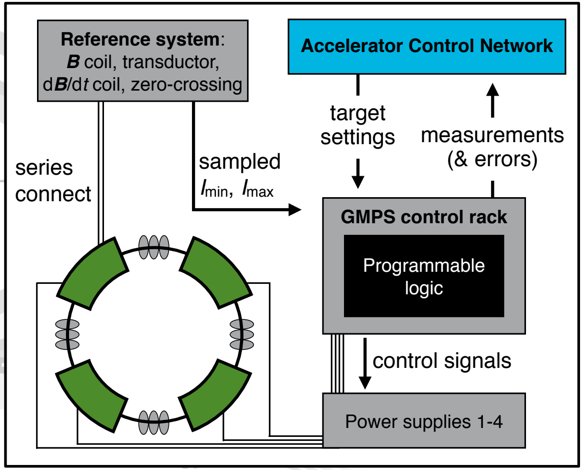
Real-time accelerator control

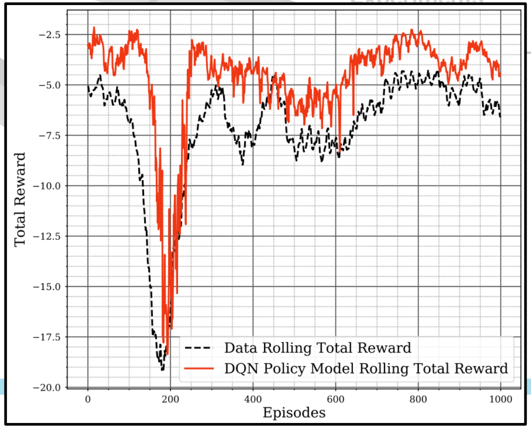


Booster GMPS

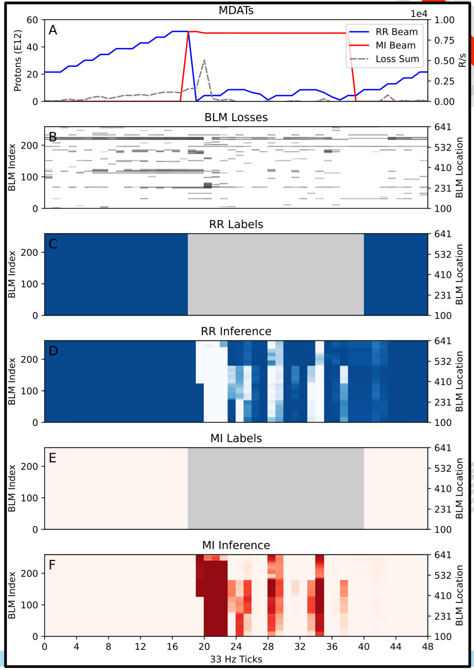
Real-time **reinforcement learning agent** in FPGA to regulate Gradient Magnet Power Supply; replace a traditional PID loop — shows improvement in reward (reduced magnet current error)

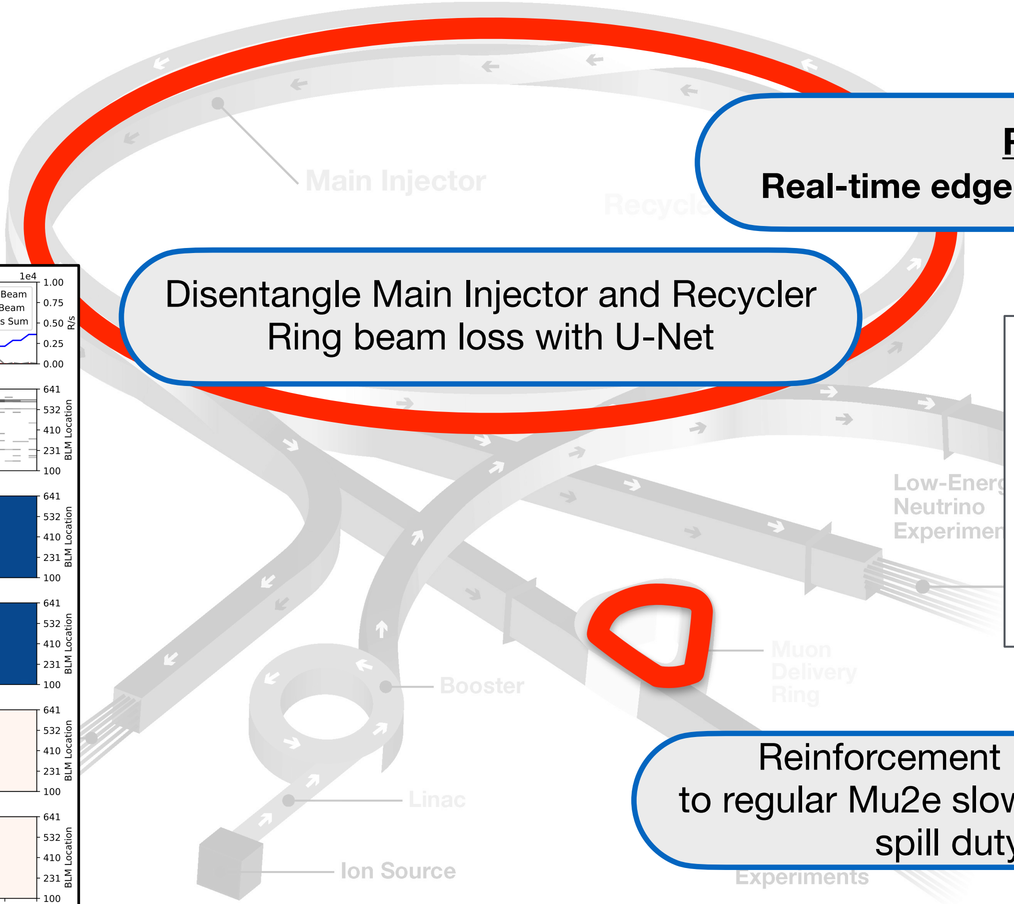
Development of digital twin for simulation framework





Real-time accelerator control



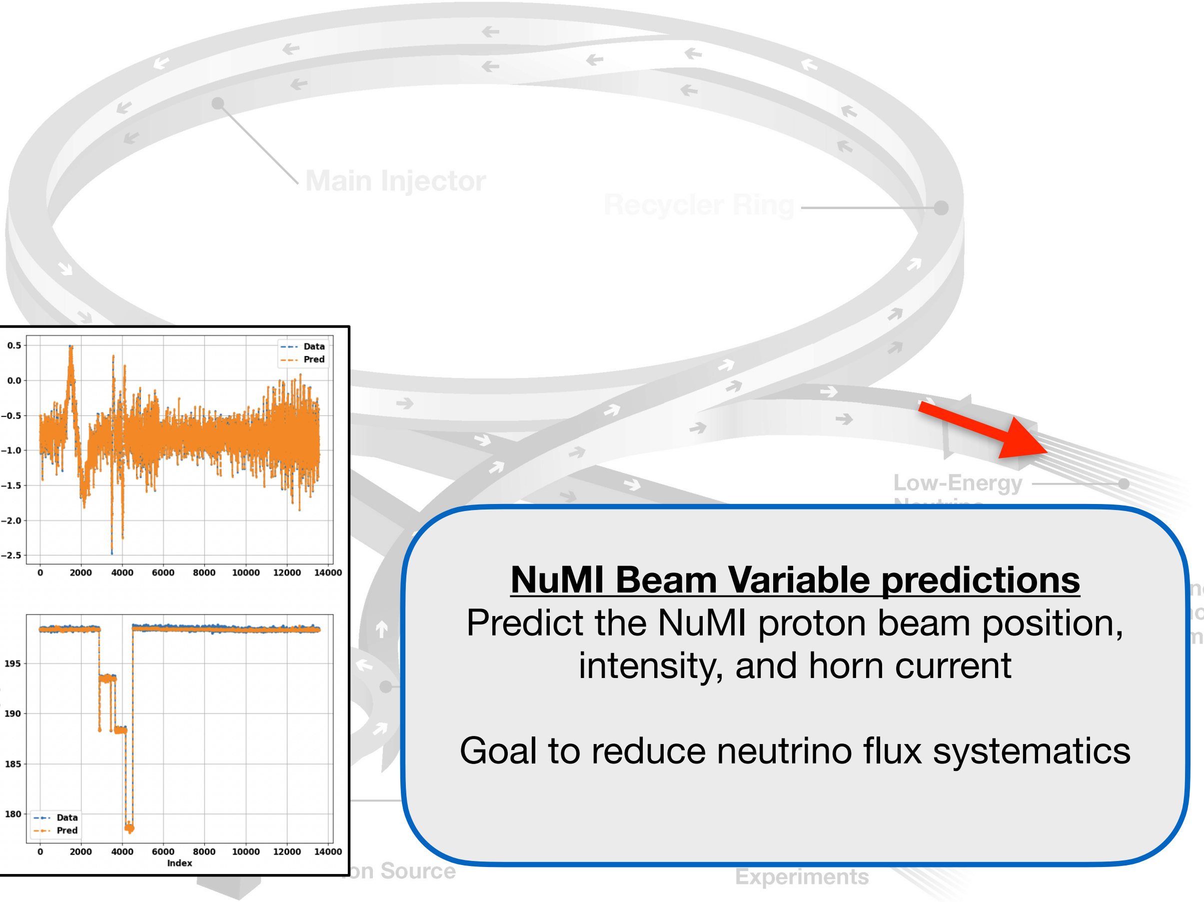


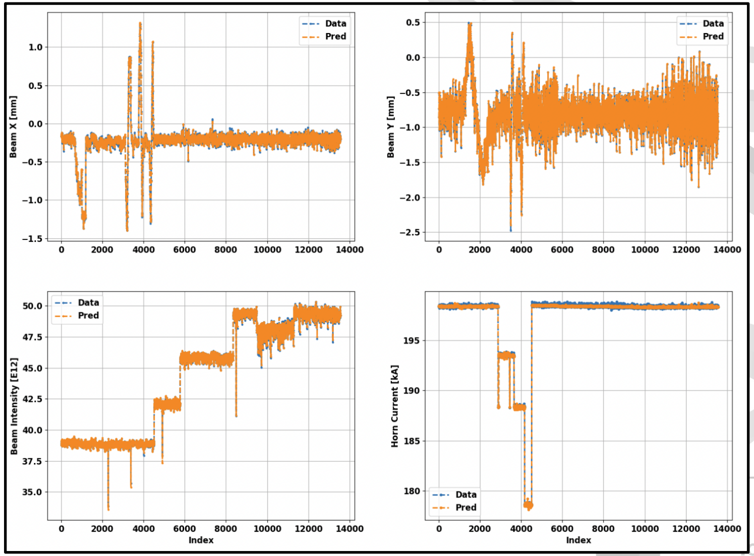
READS
Real-time edge AI distributed system

Disentangle Main Injector and Recycler Ring beam loss with U-Net

Reinforcement learning agent to regular Mu2e slow spill and increase spill duty factor

Real-time accelerator control





<u>NuMI Beam Variable predictions</u>

Predict the NuMI proton beam position, intensity, and horn current

Goal to reduce neutrino flux systematics

Fermilab

Inverse problem for nuclear modeling/tuning

Domain adaptation for robust learning (data/MC, faults,…)

Normalizing flows for accelerating MCs

Fast ML Geant/Detector simulation

Simulation based-inference

Improved reconstruction and analysis sensitivity

Algorithms for HEP science

Uncertainty quantification and fault tolerance

Operations and control systems

Geometric deep learning for HEP data representations

Physics-constrained ML; inductive bias

Anomaly detection and monitoring

Real-time AI systems at edge

Computing hardware and infrastructure

Inverse problem for nuclear modeling/tuning

Normalizing flows for accelerating MCs

Fast ML Geant/Detector simulation

Domain adaptation for robust learning (data/MC, faults,…)

Simulation based-inference

Improved reconstruction and analysis sensitivity

Algorithms for HEP science

Uncertainty quantification and fault tolerance

Operations and control systems

Geometric deep learning for HEP data representations

Physics-constrained ML; inductive bias

Distributed real-time control systems

Anomaly detection and monitoring

Real-time embedded system AI; unique engagement w/industry and other SC domains

Computationally efficient model training and implementation

Real-time AI systems at edge

Computing hardware and infrastructure

AI-on-chip: 1st detector ASIC; novel microelectronics on-sensor AI

Open-source tool flows and AI democratization; FAIR ML data and benchmarking

Inverse problem for nuclear modeling/tuning

Domain adaptation for robust learning (data/MC, faults,…)

Predictive maintenance from big data

Normalizing flows for accelerating MCs

Optimization algorithms for improved accelerator tuning

Fast ML Geant/Detector simulation

Simulation based-inference

Improved reconstruction and analysis sensitivity

Algorithms for HEP science

Uncertainty quantification and fault tolerance

Operations and control systems

Geometric deep learning for HEP data representations

Physics-constrained ML; inductive bias

Distributed real-time control systems

Self-driving, reinforcement learning systems for control and operations

Anomaly detection and monitoring

Real-time embedded system AI; unique engagement w/industry and other SC domains

Computationally efficient model training and implementation

Real-time AI systems at edge

Computing hardware and infrastructure

AI-on-chip: 1st detector ASIC; novel microelectronics on-sensor AI

Open-source tool flows and AI democratization; FAIR ML data and benchmarking

Inverse problem for nuclear modeling/tuning

Normalizing flows for accelerating MCs

Fast ML Geant/Detector simulation

Domain adaptation for robust learning (data/MC, faults,…)

Simulation based-inference

Predictive maintenance from big data

Optimization algorithms for improved accelerator tuning

Improved reconstruction and analysis sensitivity

Algorithms for HEP science

Uncertainty quantification and fault tolerance

Operations and control systems

Geometric deep learning for HEP data representations

Physics-constrained ML; inductive bias

Distributed real-time control systems

Self-driving, reinforcement learning systems for control and operations

Anomaly detection and monitoring

Real-time embedded system AI; unique engagement w/industry and other SC domains

Computationally efficient model training and implementation

Real-time AI systems at edge

Computing hardware and infrastructure

AI analysis facilities supporting multiple ML workflows

AI-on-chip: 1st detector ASIC; novel microelectronics on-sensor AI

Open-source tool flows and AI democratization; FAIR ML data and benchmarking

Integrating novel heterogeneous computing platforms for high throughput HEP workflows

Inverse problem for nuclear modeling/tuning

Normalizing flows for accelerating MCs

Fast ML Geant/Detector simulation

Domain adaptation for robust learning (data/MC, faults,…)

Simulation based-inference

Predictive maintenance from big data

Optimization algorithms for improved accelerator tuning

Improved reconstruction and analysis sensitivity

Algorithms for HEP science

Uncertainty quantification and fault tolerance

Operations and control systems

Geometric deep learning for HEP data representations

Physics-constrained ML; inductive bias

Distributed real-time control systems

Self-driving, reinforcement learning systems for control and operations

Anomaly detection and monitoring

Real-time embedded system AI; unique engagement w/industry and other SC domains

Computationally efficient model training and implementation

Real-time AI systems at edge

Computing hardware and infrastructure

AI analysis facilities supporting multiple ML workflows

AI-on-chip: 1st detector ASIC; novel microelectronics on-sensor AI

Open-source tool flows and AI democratization; FAIR ML data and benchmarking

Integrating novel heterogeneous computing platforms for high throughput HEP workflows

Huge amount activity cutting across all HEP frontiers