

# Framework-Agnostic ML Training and Inference for EICrecon

Second ePIC AI Town Hall

December 11, 2024

Dmitry Kalinkin

# Goal and scope

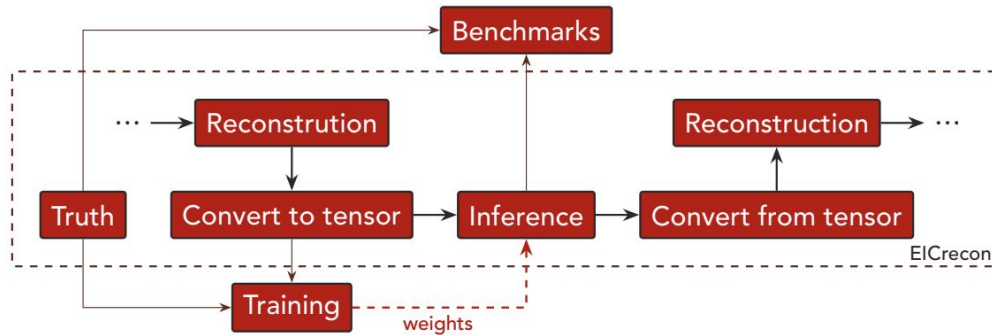
Provide a clean way to deploy AI/ML models within EICrecon reconstruction

- Move interaction with the inference engine (ONNX, TMVA, etc) into a separate reusable algorithm
- Exchange format for tensor data
  - Exchange of tensors in-memory between EICrecon algorithms
  - Export of tensors to disk for training outside of EICrecon
- Reproducible training – provide a working example of training in CI

**Ultimate goal – get more AI/ML deployed in production**

# Current status

- [Original proposal](#) for PODIO Tensor type presented in October
- ONNX backend and EEEMCal PID example implemented in ([Er#1618](#))
- Example benchmark for training and validation using CatBoost ([d\\_b#91](#))



```
edm4eic::Tensor:
Description: "Tensor type for use in training in inference of ML models"
Author: "D. Kalinkin"
Members:
- int32_t      elementType // Data type in the same encoding as "ONNXTensorElementType", ...
- int64_t      shape      // Vector of tensor lengths along its axes
- float        floatData  // Iff elementType==1, values are stored here
- int64_t      int64Data  // Iff elementType==7, values are stored here
```

- Next step is for other developers to try to apply this framework for their project
- This should be useful for the upcoming AI/ML Hackathon at the next ePIC Collaboration meeting (Jan 20-24, Frascati)