Project 1: Deep Neural Networks as electronic circuits

- Deep Neural Networks are more and more used in HEP analyses
  - main advantage is accuracy gain
  - but NN can also be a shortcut to accelerate complex algorithms (tracking, calo reco, etc) by learning how to approximate reco answers
- The challenge to use DNNs in HEP DAQ stands in the low-latency requirement (<10 ms for L1 @LHC)
- We propose to carry on an R&D to develop tools to integrate low-latency DNNs on real-time architectures
  - Training infrastructure (including model compression to reduce resources)
  - Model integration into FPGA frameworks via HLS4ML library
  - Support of all state-of-the-art architectures
  - Training opportunities to educate the community in using the tool
  - Generalization to emerging technologies (DeepLearning processors, neuromorphic chips, etc.)

REFERENCE: https://arxiv.org/pdf/1804.06913.pdf

Project 2: Deep Learning inference as a service for offline and online processing

- We propose to integrate Deep-Learning-optimized external resources into the real-time data processing workflow of HEP experiments
  - HPC sites
  - Cloud resources
- These sites (will) offer state-of-the-art computing infrastructures, optimized for DL (e.g. heterogenous platforms with high GPU/CPU rate)
- Turning our problems into DL inference, we could explore at best these resources
  - as standard processing
  - as opportunistic processing (e.g., extra trigger rate for dedicated use cases such as B physics)
- We propose to develop the communication infrastructure and demonstrators of the technologies within the LHC experiments
- With this approach, one could envision an opportunistic extension of the LHC physics reach, exploring computing resources beyond current constraints

REFERENCE: https://arxiv.org/pdf/1904.08986.pdf