Overview In the next generation of science experiments, the demands for computing resources are expected to outstrip our existing computing infrastructure's capabilities. In light of this, a radical rethinking of the cyberinfrastructure is needed. With the onset of deep learning, parallelized processing architectures have emerged as a solution. When combined with deep learning algorithms, parallelized processing architectures, such as Graphics Processing Units (GPUs) and Field Programmable Gate Arrays (FPGAs), have been shown to give large speedups in computing when compared with conventional CPUs. Based on the success of the NSF Advancing Science with Accelerated Machine Learning Harnessing the Data Revolution preparatory grant, we aim to harness these developments by bringing this technology to the scientific community and fostering an ecosystem where scientists across different domains will collaborate to solve critical problems.

This proposal harnesses the existing Fast Machine Learning community effort as a platform to expand accelerated machine learning. This Fast Machine Learning community consists of a broad range of scientists, engineers who regularly meet to discuss computing problems from many different domains. Our common theme is that we are pursuing deep learning strategies on accelerated hardware as a transformative solution for many domains. We have worked to create a common language to enable cross domain communication. At the heart of this work is the development of deep learning strategies with emerging processor technology.

By extending the community further to include astrophysics and neuroscience, we propose to construct an institute where scientific domains from a broad range of fields can share their computational challenges, construct solutions, and provide these solutions back to the community. Building on our planning grant, we aim to transition from a proof of concept to full fledged integration of accelerated ML technologies into the operation of both the Laser Gravitational Wave Observatory(LIGO) and Large Hadron Collider(LHC) experiments. Furthermore, with the addition of astrophysics, and neutrino physics, we aim to build on our LIGO developments to include deep learning based acceleration for the detection of transients and supernovae towards a deep leraning paradigm within multi messenger astronomy. With neuroscience, we aim to adapt our developments towards neural readout processing by allowing for real time processing of neural scans and fast adaptive learning to allow for enhanced data processing.

The open source tools to be developed as part of these activities will be readily shared with the broader scientific domains. With our existing codebases, we have already started this transfer of knowledge to common frameworks. The project will create an advisory group, including members of large and small projects, members of the neutrino physics, multi-messenger astronomy community, industry partners, computer scientists, and computational biologists. We will continue to support our successful workshops and boot camps where we can train students and researchers on how to use and contribute to our framework, creating a wide network of contributors and developers across key science missions. Our goal as an institute is to bring awareness to the need for hardware-based ML co-design and show that with the right knowledge, these tools can provide concrete solutions to exponentially growing data challenge needs across all science domains.

Intellectual Merit Through this proposal, we aim to create a roadmap to empower scientists to take advantage of the technological developments that are underway. We aim to accelerate the use of hardware in a broad range of fields, including high energy physics, astrophysics, and neuroscience. Furthermore, we aim to provide a toolkit and knowledge base so that the rest of the scientific community can quickly take advantage of the rapid technological developments underway.

Broader Impacts This work pushes the frontiers of deep learning at scale, demonstrating the versatility of these methods to enable science in the big data era. The students to be mentored in this research will interact closely with our industry partners, creating new career opportunities and strengthening synergies between academia and industry. Besides sharing our software tools with the community, we will continue to bring awareness to the data revolution underway.