NSF Workshop on Developing a Roadmap towards the Next Generation of Smart Cyberinfrastructure

White Paper

Delivering AI and ML capable cyberinfrastructures as a Service

Ling Liu
School of CS, Georgia Institute of Technology

Artificial Intelligence (AI) and Machine Learning (ML) have penetrated every discipline of science and engineering. Many scientific discovery and engineering breakthrough today are empowered by AI and ML capabilities. NSF is in the unique position to develop the next generation of smart Cyberinfrastructure for supporting and facilitating AI-enabled scientific research discovery and engineering innovations in all sciences and engineering disciplines.

AI and ML capable hardware cyber-infrastructures.
As AI and ML capabilities rapidly evolve in science and engineering research and development efforts supported by NSF, it is vital to scale NSF supported cyber-infrastructure from experimentation to implementation, enabling and facilitating researchers and graduate students from different science and engineering disciplines to conduct field experiments using advanced AI-capable cyber infrastructures sponsored by NSF, successfully enabling universities and research institutes to achieve AI at scale. I below list three examples

(1) Huge Data Capable AI-ML model training infrastructures, including various scales of GPU clusters with industry strength huge model training capability.
(2) Large scale Federated Learning cyber infrastructure, including large scale AI-compute clusters, each with large number AI-capable computer nodes, enabling distributed federated training implementation and experimentation.
(3) A variety of EdgeSystems, enabling EdgeAI infrastructures for large scale implementation and experimentation of Edge AI model training, model prediction and active-learning.

Taking Federated Learning Infrastructures as an example. Traditionally, training ML models requires all data to be residing in the same trusted compute server. For huge data, this can be communication intensive in addition to privacy concerns and legislations such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA). Federated deep learning has emerged as an important alternative AI-capable cyberinfrastructure and distributed AI-training paradigm. Federated Learning (FL) has been fueled by a number of big data companies, represented by Google, Facebook, Amazon, Apple. In a typical federated deep learning system, each data owner (participant) maintains its own data locally and follows a federated learning protocol where only updates of the model training parameters are shared with the trusted parameter server (model training aggregator). Participants as workers are responsible for training the same model on different mini-batches of the huge data (compute intensive tasks). In each training iteration, each participant sends its local parameter updates to the parameter server, typically hosted

in the Cloud, which stores, aggregates and maintains a set of shared parameters. The parameter server shares the aggregated parameters with each of the participants in the federated learning system and each participant updates its local parameters in the subsequent iteration. This distributed training process iterates until the model training reaches the pre-defined convergence condition. In these FL scenarios, participants are heterogeneous compute nodes that communicate with the parameter server in either a client-server style or decentralized peer to peer style. We refer to this in-network memory computing paradigm a Cloud coordinated in-network memory computing.

Setting such federated learning cyberinfrastructure at scale is not only a huge challenge for all academic researchers and graduate students and it also requires large up front capital investments. By providing the NSF-sponsored smart cyberinfrastructures to facilitate federated learning, on one hand, it will fuel the research innovation and advancement in federated learning systems, algorithms and optimizations; and on the other hand, it will enable many scientific research labs to train on their local data collections and share only parameters, which can be significantly beneficial for hospitals, healthcare professionals, virus research laboratories, and so forth. Such federated learning empowerment will shorten the path and sparkle the lights towards new, transformative scientific discovery and engineering innovations.

AI and ML capable Software Cyberinfrastructures. The next generation smart cyberinfrastructure should also be equipped with AI and ML enabled software as a service platforms and tools. Such AI software agents can be instrumental in the event of rare virus spreading. For example, if there is a shortage of doctors, caregivers can remotely monitor at risk patients and robots can deliver medicines for humans to minimize infection spreading and help reduce reliance on doctors for routine readings and so forth.

In the digital transformation and big data age, AI is an important tool, which will make changes in the ways how scientists and engineers tackle problems in every area of our sciences and engineering and daily life. AI-capable software agents will make things easier for scientists and engineers to focus more closely on innovation and technology trends.

AI and ML Security and Privacy Cyberinfrastructures. AI models and ML models are vulnerable to adversarial perturbation attacks. A smart cyberinfrastructure should be capable of providing built-in secure AI workflows throughout the AI-training and AI prediction workflow life cycle, enabling data input guard, AI inference guards and AI output guards.

I have done research in the above areas, especially in federated learning infrastructures and secure AI/ML workflow system infrastructures. I am happy to share my experiences, expertise and lessons learned with the workshop attendees and contribute to discussions at the workshop and learn from other participants on their visions and experiences on the open challenges in building the next generation smart cyber-infrastructures at the workshop.