

Title: Quest for Scaling Earthly AI Holistically - Advancing AI to help children with speech and language service needs

Abstract: AI has become an increasingly powerful technology force that promises to impact all aspects of our society, from transportation to healthcare and education to sustainability. But the diverse layers of software abstractions, hardware heterogeneity, and data privacy concerns have made the development of optimized AI solutions extremely challenging. This results in the business world's expensive investment only on a handful of selective and "profitable" AI solutions, leaving many critical societal needs, such as equitable education and sustainability, much less addressed than deserved. For example, nearly 3.4 million children in the U.S. require speech and language services under the Individuals with Disabilities Education Act (IDEA) and are at risk of falling behind in their academic and social-emotional development without timely intervention by Speech and Language Pathologists (SLPs). Unfortunately, there is a significant shortage of SLPs and the COVID pandemic has further exacerbated this gap, making it almost impossible for SLPs to provide individualized services for children.

Through the recently established NSF National AI Institute for Exceptional Education, we envision a transformative approach to address this challenge. We aim to develop advanced AI technologies to scale SLPs' availability and services such that no child in need of speech and language services is left behind. Towards this end, we propose to develop two novel AI solutions: (1) AI Screener to enable universal early screening for all children, and (2) AI Orchestrator to work with SLPs and teachers to provide individualized interventions for children with their formal Individualized Educational Plan (IEP). In developing these solutions, we will advance foundational AI technologies, enhance our understanding of children's speech and language development, serve as a nexus point for all special education stakeholders, and represent a fundamental paradigm shift in how SLPs serve children in need of ability-based speech and language services.

In this talk, I will discuss the rationale behind the Institute's vision, the technical approaches we plan to take, and the corresponding research challenges we must overcome. I will contextualize these efforts in an ultimate research goal of transforming the current AI innovation ecosystem to truly democratize AI for a better society.

第二個演講主題及摘要如下：

Title: Make every bit of bandwidth count: AI systems innovation for accelerating AI in the era of Cloud Computing

Abstract: The world of computing has seen significant changes in the increasing adoption of AI and its wide deployment onto the cloud. The computation patterns driven by modern machine learning models (such as deep neural networks, transformers, large-language models) often require accessing large volumes of data, which poses significant challenges for both programming and the underlying computing infrastructures. On the one hand, the programming needs to manage the complexity of moving data across different storage hierarchy (including remote storage nodes). On the other hand, the computing infrastructure keeps getting more complex with each upgrade: larger memory capacity, faster data access speed, and higher data I/O bandwidth. In this talk, we argue that many such design options are, however, not necessarily best suited for existing AI workloads' acceleration. In fact, we demonstrate that there are still ample opportunities for AI workload acceleration on existing computing infrastructure, provided we can design effective AI system tools to characterize the system bottlenecks, and novel hardware and software techniques to fully utilize the given system's I/O bandwidth. This opens a set of new design optimization options for acceleration AI workloads on the cloud.