



## **The Era of AI Hardware**

**Myung-Hee Na, Ph. D., Distinguished Engineer, IBM Research**

Hardware has so far played a supporting part in the development of narrow AI. However, the adoption in broad AI calls for increasing the focus on innovation of hardware to address challenges such as bandwidth, latency and scalability requirements of complex AI models.

The concurrent evolution of broad AI with purpose-built hardware will shift the traditional balances between cloud and edge, structured and unstructured data, and training and inference. Heterogeneous system architectures are already being effectively developed where varied compute resources, including high-bandwidth CPUs, specialized AI accelerators and high-performance networking are infused in each node to yield significant performance improvement.

Looking beyond present needs and capabilities to accelerate AI, we believe that purpose-built hardware solutions are required to unlock the projected exponential gains in AI computations. In this talk, we will share our vision of a roadmap of specialized technologies to enable an era of AI hardware, including reduced precision approaches with Digital AI Cores, non-Von-Neumann approaches with Analog AI Cores, and the emergence of quantum computing solutions for AI workloads.