**Abstract:**

Training algorithms for large language models are often communication heavy. As a result, these models are trained dominantly in a centralized environment such as data centers with fast network connections. This strong dependency on fast interconnections is becoming the limiting factor of further scaling for the data center setting and alternative decentralized infrastructures such as spot instances and geo-distributed volunteer computes. In this talk, I will discuss our research in communication-efficient distributed learning and our current effort in training foundation models in a decentralized way.

**About the Speaker:**

Binhang Yuan is an Assistant Professor at the Department of Computer Science and Engineering (CSE), the Hong Kong University of Science and Technology (HKUST). He received his Ph.D. and master’s degrees from Rice University and his bachelor’s degree from Fudan University. Before joining HKUST, he was a Postdoc at the Swiss Federal Institute of Technology Zurich (ETH Zurich). His main research interests are in data management systems for machine learning, distributed and decentralized machine learning systems.