Seminar

ECU COMPUTER SCIENCE

Building Distributed Machine Learning Supercomputers

Abstract. This presentation explores the architectural foundation of modern ML compute infrastructure. We begin with a quick refresher on Graphics Processing Units **(GPUs)** and then a technical introduction to the Tensor Processing Unit **(TPU)**, examining its specialized role in accelerating large-scale training and inference. Finally, we'll detail some essential distributed strategies and frameworks required to efficiently manage and orchestrate GPUs and TPUs into an effective ML supercomputer.

Biography. Mr. Saarenvirta is a Staff Technical Solutions Consultant at Google Cloud who spends most of his time helping Google customers build state-of-the-art ML training and inference clusters on Google Cloud Platform (GCP)

Mr. Erik Saarenvirta
Google | AI Infrastructure

https://www.linkedin.com/in/erik-saarenvirta/

Friday November 21, 2025
Time: 2:00 – 3:00pm

Microsoft Teams
Join the meeting now
Meeting ID: 279 685 937 875 66
Passcode: f7sW7Fv3

Contact: Dr. Kamran Sartipi
Dept. of Computer Science, ECU
https://cs.ecu.edu/sartipik16/CSseminar/