

COMPUTER SCIENCE  
SEMINAR

*Data Collection for Data-Centric AI*

Dr. Fatemeh Nargesian  
University of Rochester

**Abstract:** Abstract: The holy grail of data-centric AI is to collect high-quality labeled data sets for the purpose of training ML models. Data collection has become an active area of research in the data management community due to the importance of handling large amounts of training data. This talk will examine the data collection techniques that can be used to discover, augment, or generate datasets from existing data lakes. I will also cover data tailoring that is to ensure that the collected data set for analysis has an appropriate representation of relevant (demographic) groups: it meets desired distribution requirements. I will conclude by introducing some of the interesting research challenges that remain in the data collection landscape.

Biography: Fatemeh Nargesian is an assistant professor in the Department of Computer Science, at the University of Rochester. She got her PhD at the University of Toronto and was a research intern at IBM Watson. Before the University of Toronto, she worked at Clinical Health and Informatics Group at McGill University. Her primary research interests are in data intelligence focused on data discovery, data, integration, and data for ML.

**\*\*Join Zoom Meeting\*\***

Venue: <https://emory.zoom.us/j/98352727203>

Friday, October 22, 2021, 1:00 pm  
<https://emory.zoom.us/j/98352727203>

COMPUTER SCIENCE  
EMORY UNIVERSITY