
THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Faculty Seminar

Room 1 ,Ziskind Building
on Wednesday, Jan 05, 2022at 11:15

Amir GiladDuke University

Informed Data Science

Abstract:

Data science has become prevalent in various fields that affect day-to-day lives, such as healthcare, banking, and the job market. The process of developing data science applications usually consists of several automatic systems that manipulate and prepare the data in different manners. Examples of automatic data manipulations and preparations include generating synthetic data, interactive data exploration, repairing the data, and labeling it for machine learning. These systems can be highly complex and even data scientists can find it difficult to understand and verify their output. Moreover, uninformed use of these systems can lead to errors that may affect the quality of the results of such applications. In the talk, I will highlight prominent challenges in the data science process and present three aspects of my research that are meant to assist in this process. In particular, I will present a solution that generates natural language explanations for query results, a tool for generating synthetic linked data, and a solution that explains complex queries using abstract database instances. Zoom: <https://weizmann.zoom.us/j/97604026721?pwd=SVZnd3NsQXBqNHFhOU04bTFNQVRhQT09>