THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Faculty Seminar

Room 155 ,Ziskind Building
on Wednesday, Aug 17, 2022at 10:30

Efrat ShimronUC Berkeley

Data Crimes: The Risk in Naive Training of Medical AI Algorithms

Abstract:

Although open-access databases are an important resource in the current deep learning (DL) era, they are sometimes used in an "off label" manner: data published for one task are used during training of algorithms for a different task. In this seminar I will show that this leads to biased, overly optimistic results of well-known inverse problem solvers, focusing on algorithms developed for magnetic resonance imaging (MRI) reconstruction. I will show that when such algorithms are trained using off-label data, they yield biased results, with up to 48% artificial improvement. The underlying cause is that public databases are often preprocessed using hidden pipelines, which change the data features and improve the inverse problem conditioning. My work shows that canonical algorithms - Compressed Sensing, Dictionary Learning, and Deep Learning algorithms - are all prone to this form of bias. Furthermore, once trained, these algorithms exhibit poor generalization to real-world data, thus they could produce unreliable results in clinical setups. To raise awareness to the growing problem of naive use of public data and the associated biased results, the term "data crimes" is coined.