

## SYSEN 8100-1: Systems Seminar Series

### Avigdor Gal: Loch Data and Other Monsters on Creating Data Ecosystems, The Intelligent Way

Nick Kunz [NetID: [nhk37](#)] [nhk37@cornell.edu](mailto:nhk37@cornell.edu)

October 12, 2022

On October 7, 2022, Dr. Avigdor Gal, the Benjamin and Florence Free Chaired Professor of Data Science at the Faculty of Industrial Engineering & Management at Technion - Israel Institute of Technology, presented a case for data integration and schema matching in the context of human in the loop algorithms and testing. The focus of his presentation was largely to demonstrate state-of-the-art methods for identifying expert decision making abilities in that regard.

Dr. Gal first described data ecosystems and what he meant by a “*data monster*”. A “*data monster*” was a direct reference to the overwhelming abundance of data, specifically regarding the amount of data sources themselves and how “*data monsters*” emerge through the process of data integration. He also mentioned important topics regarding data fairness, trust, completeness, accuracy, and consistency, as important considerations across domains.

Dr. Gal went on to compare the decision making abilities of humans to algorithms and the link between them. He introduced a novel schema matching algorithm, Matching Expert Identification (MEXI), a multi-class, multi-label classification algorithm for identifying expert decision making. MEXI’s ability to identify expertise was not limited to schema matching, but was also applicable in broader and more generalized tasks, although those details were beyond the scope of the presentation.

It was interesting to learn the performance metrics used to measure the broader notion of expertise. In the case of schema matching, those performance metrics were: 1) precision, 2) thorough, 3) correlated, and 4) calibrated. Dr. Gal was careful to make the distinction between correlated vs. calibrated, where correlation means higher confidence to good results, and calibrated means the degree of confidence in those results. Dr. Avigdor raises interesting questions about the exactness of what we mean when we make claims of expertise, in both human and machine intelligence.