NJIT | INSTITUTE FOR DATA SCIENCE

Department of Data Science

Data Science Seminar Series

Towards Explainable State-of-the-Art Artificial Intelligence Models

Sarah Adel Bargal, Ph.D.
Research Assistant Professor
Boston University

Date: Thursday, February 10th, 2022
Time: 2:30 PM – 3:30 PM EST
Location: Zoom Virtual Room
Web Link: Zoom Meeting Room Link

Deep learning is now widely used in state-of-the-art Artificial Intelligence (AI) technology. A Deep Neural Network (DNN) model however is, thus far, a “black box.” AI applications in finance, medicine, and autonomous vehicles demand accurate and justifiable predictions, barring most deep learning methods from use. Understanding what is going on inside the “black box” of a DNN, what the model has learned, and how the training data influenced that learning are all instrumental as AI serves humans and should be accountable to humans and society. In this talk, I will present work where we (1) visualize cues that contribute to a deep model’s classification/captioning output using the model’s internal representation, (2) demonstrate that such cues can be used to localize regions/segments that correspond with a specific action, or phrase from a caption, without explicitly optimizing/training for these tasks, and (3) propose frameworks that utilize such cues to train better models and frameworks that are designed to be inherently more explainable.

Sarah is a Research Assistant Professor in the Department of Computer Science of Boston University, Co-director of AI4ALL at Boston University, and a member of the Image and Video Computing (IVC) Group. Sarah first joined the IVC Group in 2013 as a PhD student working with Stan Sclaroff, and later as a Postdoctoral Associate working with Stan Sclaroff and Kate Saenko. In 2019, she received a Ph.D. in Computer Science from Boston University. She is a recipient of the IBM PhD Fellowship, Hariri Graduate Fellowship, Outstanding Teaching Fellow Award, among other recognitions. Sarah is currently serving as a guest editor for a special issue of the Frontiers in Computer Science Journal. Previously, Sarah received a M.Sc. in Computer Science from the American University in Cairo, after which she served as a lecturer of Computer Science and Mathematics at the Gulf University for Science and Technology. Sarah received her B.Sc. in Computer Science from Kuwait University. Sarah’s research interests are in machine learning, computer vision, and explainable artificial intelligence, with a current focus on making artificial intelligence systems explainable and accountable to humans and society.