Dr. Adnan Darwiche

Explaining and Verifying AI Systems

Explaining the decisions of AI systems and formally verifying their properties have come into focus recently. In this talk, Dr. Darwiche will discuss an approach for explaining and verifying Bayesian network classifiers, which is based on compiling them into equivalent and symbolic decision graphs. He will also discuss a new class of circuits that are as expressive as neural networks and that can be synthesized from Bayesian network models, allowing one to provide formal guarantees on their behaviors regardless of how they are trained from data.

November 7th, 4-5 p.m., THH 301

This lecture satisfies requirements for

CSCI 591: Research Colloquium.

Dr. Adnan Darwiche is a professor and chairman of the computer science department at UCLA. He directs the automated reasoning group which focuses on probabilistic and logical reasoning, and their applications including to machine learning (http://reasoning.cs.ucla.edu/).