Abstract:

In this talk, I will present two novel, complementary methods for making deep learning models more reliable and interpretable. First, I will present AI2, a technique for proving that a neural network satisfies a given property (e.g., robustness). The key idea is to leverage abstract interpretation to soundly over-approximate the network's behavior, enabling the analysis of large convolutional models, for the first time. Then, I will present DL2, a method which bridges logic and differentiable reasoning, allowing one to both pose interesting queries on deep models as well as train them to satisfy formal properties which capture domain knowledge.

Finally, I will briefly discuss few promising research directions where applications of automated reasoning, both discrete and probabilistic, can be effective in solving key challenges in systems (e.g., programmable networks) and security (e.g., differential privacy).

Bio: Dana Drachsler-Cohen is an ETH Postdoctoral Fellow at the department of Computer Science, ETH Zurich. Her research interests span automated reasoning, program synthesis and machine learning. She obtained her PhD from the Computer Science Department at the Technion in 2017.