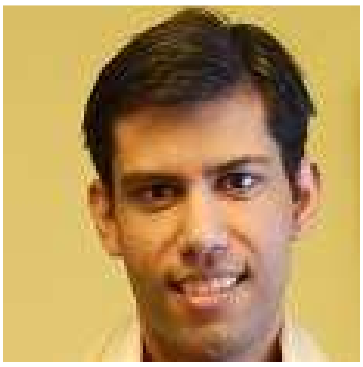
UNIVERSITY AT ALBANY
State University of New York

**College of Engineering and Applied Sciences**
**Computer Science Department**

# Colloquium



*Date, time & venue*: Monday, November 7th, 2016, 11:00-12:15 in BB129

*Title*: **On Making Machine Learning Safe**

*Speaker*: **Kush R. Varshney, Research Staff Member, IBM T. J. Watson Research Center in Yorktown Heights, NY**
**(http://researcher.watson.ibm.com/researcher/view.php?person=us-krvarshn)**

**Abstract:** Machine learning algorithms are increasingly influencing our decisions and interacting with us in all parts of our daily lives. Therefore, just like for chemical plants, roads, vehicles, and myriad other systems, we must ensure that systems involving machine learning are safe. In this talk, we first discuss the definition of safety in terms of risk, epistemic uncertainty, and the harm incurred by unwanted outcomes. Then we examine dimensions along which certain real-world applications may not be completely amenable to the foundational principle of modern statistical machine learning: empirical risk minimization. In particular, we note an emerging dichotomy of applications: ones in which safety is important and risk minimization is not the complete story (we name these Type A applications), and ones in which safety is not so critical and risk minimization is sufficient (we name these Type B applications). Then, we discuss how four different strategies for achieving safety in engineering (inherently safe design, safety reserves, safe fail, and procedural safeguards) can be mapped to the machine learning context through interpretability and causality of predictive models, objectives beyond expected prediction accuracy, human involvement for labeling difficult or rare examples, and user experience design of software. Finally, we detail principled optimization-based formulations for learning Boolean rule-based classifiers that are interpretable and therefore provide inherently safe design.

**Speaker's brief bio:** Kush R. Varshney is a research staff member in the Mathematical Sciences Department at the IBM T. J. Watson Research Center in Yorktown Heights, NY. He received the Ph. D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2010. He applies data science and predictive analytics to human capital management, healthcare, olfaction, public affairs, and international development. He conducts academic research on the theory and methods of statistical signal processing and machine learning. His work has been recognized through best paper awards at the Fusion 2009, SOLI 2013, KDD 2014, and SDM 2015 conferences. Dr. Varshney is co-director of the IBM Social Good Fellowship program.