Recent discussion in the public sphere about classification by algorithms has involved tension between competing notions of what it means for such a classification to be fair to different groups. We consider several of the key fairness conditions that lie at the heart of these debates, and discuss recent research establishing inherent trade-offs between these conditions. We also explore how the complexity of a classification rule interacts with its fairness properties, showing how natural ways of approximating a classifier via a simpler rule can act in conflict with fairness goals.

The talk will be based on joint work with Jens Ludwig, Sendhil Mullainathan, Manish Raghavan, and Cass Sunstein.

Bio sketch: Jon Kleinberg is the Tisch University Professor in the Departments of Computer Science and Information Science at Cornell University. His research focuses on the interaction of algorithms and networks, the roles they play in large-scale social and information systems, and their broader societal implications. He is a member of the National Academy of Sciences and the National Academy of Engineering, and the recipient of MacArthur, Packard, Simons, Sloan, and Vannevar Bush research fellowships, as well awards including the Harvey Prize, the Nevanlinna Prize, and the ACM Prize in Computing.