Increasingly, decisions and actions affecting people's lives are determined by automated systems processing personal data. Excitement about these systems has been accompanied by serious concerns about their opacity and the threats that they pose to privacy, fairness, and other values. Recognizing these concerns, it is important to make real-world automated decision-making systems accountable for privacy and fairness by enabling them to detect and explain violations of these values. System maintainers may leverage such accounts to repair systems to avoid future violations with minimal impact on the utility goals.

In this thesis, I aim to develop theories and tools for analyzing information use that enable practical accountability mechanisms and ensure data-driven systems respect meaningful privacy and fairness properties. In particular, I focus on two forms of information use: (i) explicit use, the direct causal influence of information, and (ii) proxy use, the indirect use of information through associations. In prior work, I have developed methods for detecting and explaining explicit information use. In this proposal, I will address issues due to covariate shifts in causal testing of machine-learned systems. Further, I will focus on a new formalization of proxy use, and tools for its detection and repair. Finally, I will explore theories of use privacy and proxy non-discrimination built on top of this formalization of proxy use.

Thesis Committee:
Anupam Datta, Chair
Jaime Carbonell
Matt Fredrikson
Sriram Rajamani, Microsoft Research
Jeannette Wing, Microsoft Research

Thesis Summary: www.cs.cmu.edu/~shayaks/sen-proposal.pdf