

Shomir Wilson (Penn State)

Title: Natural Language Processing for Privacy and Social Good

Abstract: Internet users care about their privacy, but persistent gaps exist in their understanding of companies' data practices, and more generally, how data about them can be collected and used. I will describe a trajectory of research to automatically extract information from the text of websites' and apps' posted privacy policies and to present it to consumers, regulators, and privacy researchers in ways that better respond to their needs. A recent part of this work is PrivaSeer (<https://privaseer.ist.psu.edu/>) a search engine and corpus that together make a collection of over 1M website privacy policies available and explorable for privacy stakeholders. I will also describe several other projects to apply natural language processing (NLP) to problems in privacy. I will conclude with some thoughts on the potential for NLP to help with a broad set of problems involving large volumes of legal text and the ability of consumers to make informed decisions.

Bio: Shomir Wilson is an Assistant Professor in the College of Information Sciences and Technology at the Pennsylvania State University, where he leads the Human Language Technologies Lab. His research interests span natural language processing, privacy, and computational social science. He is particularly interested in breaking down technology's "walls of text", situations where a human reader is expected to consume a large quantity of text to take action while lacking time or expertise to properly understand it. He holds \$1.4M in active grants from the National Science Foundation and the National Institutes of Health, covering research on privacy policies, usable privacy, and fairness in law enforcement. Prior to becoming faculty he held postdoctoral positions in Carnegie Mellon University's School of Computer Science and the University of Edinburgh's School of Informatics. He received his Ph.D. in Computer Science from the University of Maryland in 2011. For more about his work, visit <https://shomir.net/>.