AI for privacy abstract
According to a Deloitte assessment in 2017, involving 2000 users, “91 percent of users willingly accept legal terms and conditions without reading them before installing apps, registering Wi-Fi hotspots, accepting updates, and signing on to online services such as video streaming.” That poses a threat to the privacy of users. It is unrealistic to expect that humans will read all the privacy policies of the applications they use. Artificial Intelligence can simplify this problem and provide a realistic option to protect privacy of users by automatically identifying the information that a user is disclosing when accepting a policy. To train a model to automatically identify that information, a dataset of 92 privacy policies of applications was constructed manually. Different classes for types of information were defined. Then, strings that belong to one of the defined classes in the dataset were manually annotated. State of the art algorithms for Named Entity Recognition (NER) were trained.

The best performing algorithm for this task was “End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF” by Xuezhe Ma and Eduard Hovy, using the implementation by Guillaume Genthial. Initially f2 score of 60%. However, after error analysis were found correct results that were counted as incorrect because of diverging boundary in the string. If they captured the correct meaning, they were counted as correct, to obtain an f2 score of 82%. The majority of the remaining of the errors found in error analysis seem to be uncommon cases that should be fixed with more training data. We believe these results are promising to invest in implementing a tool that allows a user to configure their privacy preferences once and deny or accept privacy policies automatically.

Acknowledgments to Travis Breaux and Jaspreet Bhatia for their leadership.