UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Seminar

Friday December 20, 2019 at 11:30AM in Reynolds, Room 3324

Automated Generation of Privacy Policy Using Deep Models

Nastaran Bateni

Advisor: Dr. Rozita Dara
Advisory Committee: Dr. Stacey Scott

ABSTRACT:

Privacy policies are statements about how websites, applications, and any other service providers collect, use, share and manage users' data. Contents of privacy policies depend on countries laws and cross-border rules in addition to the type of service. Nowadays, the contents of privacy policies have been affected by different regulations such as the General Data Protection Regulation (GDPR). GDPR is a framework that enforces protection of personal data. It requires privacy policies to be more transparent and clear for readers. Since GDPR has come to effect, many websites and applications need to change their content complying with GDPR requirements. However, we have a limited understanding on how these legislations have impacted the content of privacy policies in recent years. Therefore, it is desirable to understand more about the evolution of the content.

This study examines the impact of GDPR on the content of privacy policies and data practices, to evaluate best practices used in creating them, and to utilize machine learning methods for automatic generation of privacy policies ensuring compliance with standards provided by GDPR or other fair information practices.

To this end, 4000 online privacy policies (written after GDPR and before) are collected from a number of different websites. They are then compared in keywords, topics and data practices using Natural Language Processing (NLP) methods. Furthermore, we categorized keyphrases into data practices such as "sharing data", "collecting data", "choice of data", "security of data", etc. As the next step, these documents will be used to train deep neural networks including CNN and LSTM to generate legal data practices and content.