THEME: Cyberspace, Privacy and Data Protection

PROJECT: Technological Solutions for Fair Unsupervised AI

Lead Supervisor: Dr Deepak Padmanabhan

Supervisory Team: Dr Deepak Padmanabhan, Dr Muiris MacCarthaigh

Primary Location: CSIT

In the digital era, governments and public organizations collect vast amounts of citizen, entity, and system data. These organizations are increasingly considering ways to use AI to inform policy and operational decisions through supervised machine learning and AI, raising important fairness and transparency questions. Yet an increasing amount of this data is unlabelled, necessitating unsupervised learning, particularly when increasingly large amounts of data are collected automatically through devices such as surveillance cameras and smart sensors. The latter trend has been facilitated by an expansion of the methods for 'passive' data collection, where data is collected through safety/surveillance cameras, IoT devices as part of smart city infrastructure, various kinds of sensors, smartphone apps, medical wearables, traffic sensing devices, public wifi and even social media monitoring.

The scope of passive data collection has been expanding, fuelled by public-private partnerships such as those with camera manufacturers and urban analytics solution providers. This project will focus on developing a suite of algorithms that embed notions of fairness within their formulation, which will ensure that the results from those algorithms would be fair. Fairness will be quantified as distributional parity along sub-groups of data determined by way of dimensions that are deemed sensitive, such as gender, ethnicity, age, location, employment and education.

The purpose of this PhD project will be:

a) to identify fairness pitfalls while employing algorithmic solutions for processing vast amounts of unlabelled data using conventional exploratory analytics algorithms,
b) to devise algorithmic techniques to embed notions of fairness within exploratory analytics algorithms such as those for clustering, retrieval, representation learning and outlier detection,
c) to benchmark and evaluate the fairness in the results of the analytics algorithms, and evaluate their appropriateness within concrete use cases within the public sector.

The project will primarily involve technology development, but will be informed and guided by the politics and philosophy of fairness, and more broadly the role of ethics in public service delivery and democratic government. Therefore, as well as demonstrating analytical capacity from a computer science background, the successful candidate will be expected to engage with the social sciences and where possible applications should identify knowledge of theories and concepts relevant to the study of fairness.

Primary Academic Discipline: Computer Science