Background and rationale

Data science is an interdisciplinary scientific approach that provides methods to understand and solve problems in an evidence-based manner, using data and experience. Despite the clear benefits from the adoption of data science, many firms face challenges, be that legal, organisational, or business practices, when seeking to implement and embed data science within their existing frameworks. The rationale for this workshop was to bring together data science practitioners to share their experiences with, and insights on these challenges.

Workshop format

In this workshop, panel and audience members drew on their experiences to elaborate on the data science challenges they had encountered. Many of the comments were able to be categorised within three themes: legal, organisational, or business practices. Panel and audience members were from business, academia, and think-tanks. In the first half of the workshop the panel members discussed their own experience from prepared remarks and in the second half of the workshop audience members commented and responded, initially via a smartphone app, based on their own experience.

Organisers

The organisers of the panel were: Rohan Alexander, Kelly Lyons, Michelle Alexopoulos, and Lisa Austin. The organisers put together a panel comprised of four speakers: Aije Egwaikhide, IBM; Inmar Givoni, Uber Advanced Technologies Group (ATG); Ofer Shai, Omnia AI; and Aaron Shull, The Centre for International Governance Innovation (CIGI).

Egwaikhide is a Senior Data Scientist at IBM with experience in oil and gas as well as cognitive systems. She is focused on actionable insights using data science techniques and improving processes through building machine learning solutions.

Givoni is a Senior Engineering Manager at Uber ATG Toronto where she builds self-driving vehicles using cutting-edge deep-learning models. In addition, she fairly regularly gives presentations to high school and university students, researchers, and non-technical audiences about machine learning, algorithms, big data, and their applications. She also participates in various career panels and career mentoring events.

Shai is the Chief AI Officer at Deloitte Canada, where he leads the IP factory. Ofer has 15 years of experience in natural language processing, predictive and advanced analytics, recommendation systems, information retrieval, and computational biology. He holds a PhD from the University of Toronto Machine Learning Group, and has held leadership positions at Kobo, Upsight, Meta, and the Chan-Zuckerberg Initiative.

Shull is Managing Director and General Counsel at CIGI. As CIGI's managing director and general counsel, he is a strategic liaison between CIGI's research programs and other departments while managing CIGI's legal affairs and advising senior management on a range of legal, operational and policy matters

Outcomes

In the first half of the workshop the panelists discussed the challenges they had observed and/or faced when adopting and using data science within an existing framework. Each panelist spoke from prepared remarks. Many points were discussed, including some of which are detailed below, which also draws upon related literature.

Shai discussed the Deloitte Survey of data science, which found a 16 per cent adoption in Canada, within industry. There were four themes from Deloitte's survey:

1) A lack of understanding
2) Lack of a trust, for instance, data breaches and misuses of data.
3) A fear of risk, for instance as fear that small start-ups may not survive for very long and so not be worth investing in a relationship with.
4) Inability to scale, for instance unsure how to integrate data scientists across an organisation.

Egwaikhide discussed aspects from her consulting experience. For instance, it can be difficult to make it clear what machine learning and AI can even do. There can be a misunderstanding between what clients think can be done and what can actually be done. Often the underlying datasets are of poor quality in terms of having irrelevant features, and insufficient quantity. Challenges with the data science as well including inappropriate use of algorithms; overfitting the training data; using fancy algorithms for simple tasks; and underfitting training data (model is too simple).

Givoni spoke about the difficulty of even defining AI, in order to better understand adoption rates. She spoke a lot about the realities of implementing these systems, pointing out that even if self-driving were 10 per cent better than a human, then a lot of people would still die. It is only when they become substantially better than humans that there would be widespread appreciation of the benefits. Furthermore, she discussed the difficulties of testing much of machine learning code as it is not testable in the same way that regular software is testable.

Finally, Shull discussed the antiquated nature of the privacy laws, and their consent-based nature which are rarely read. He went on to describe how AI and machine learning projects are commonly treated as just another project, rather than requiring large-scale thinking. Finally Shull discussed issues around finding enough skilled people.

After these individual presentations, the panel discussed various aspects, including, why has data science developed a bad reputation, the evolution of the regulatory framework, the changing labor force, and a lack of trust. Finally, there were comments from the audience. This included a discussion of the definition of data science and related terms, how to handle bad data, and the gap is between practise and marketing.