Avoiding Algorithmic Bias

David Weinkauf, Ph.D.
Senior Policy and Technology Advisor
Information and Privacy Commissioner of Ontario

Information and Privacy Commissioner of Ontario
Commissaire à l'information et à la protection de la vie privée de l'Ontario

DAAC Summit

Olser Hoskin & Harcourt

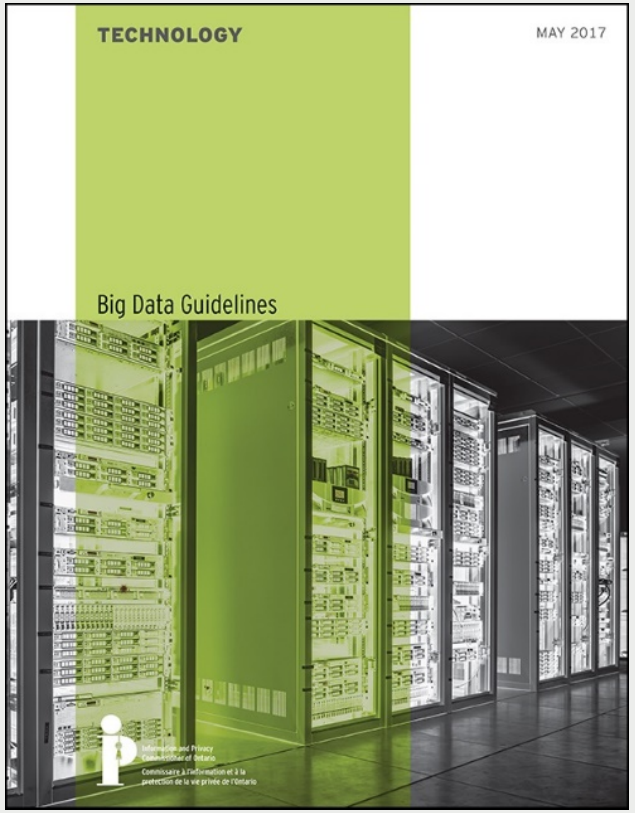
September 28, 2018

Outline

- Ontario IPC's "Big Data Guidelines"
- Bias and Canadian Privacy Laws
- Sources of bias in algorithms:
  - Existing biases in society
  - Training data
  - Design decisions made by practitioners
- Conclusion

Ontario IPC's Big Data Guidelines

- IPC released "Big Data Guidelines" in May 2017
- Designed to inform government institutions of **key issues, best practices** when conducting big data projects involving personal information
- Divides big data into **four stages**:
  1. collection
  2. integration
  3. analysis
  4. profiling
- Each stage raises a number of concerns (14 total), including bias and accuracy



Bias and Canadian Privacy Laws

- 2018 Supreme Court of Canada decision *Ewert v. Canada*
  - Mr. Ewert, who identifies as Métis, challenged Correctional Service Canada's (CSC's) use of **psychological and actuarial risk-assessment tools**
  - Argued they were developed and tested on **predominantly non-Indigenous populations**
  - SCC upheld decision finding that CSC had breached its obligations under s. 24(1) of the *Corrections and Conditional Release Act* to "take all reasonable steps to ensure that any information about an offender that it uses is as **accurate, up to date and complete** as possible"
- Principle 4.6 of Schedule 1 of the *Personal Information Protection and Electronic Documents Act* (PIPEDA) provides that "Personal information shall be as **accurate, complete, and up-to-date** as is necessary for the purposes for which it is to be used"

Existing Biases in Society

- Researchers found that a word2vec machine learning (ML) model trained on Google News articles **outputted stereotypes** such as:
  - "Man is to computer programmer as woman is to homemaker"
  - "Father is to doctor as mother is to nurse"
- The researchers then hired **human raters** to identify biases as a way of eliminating them in the model
- Until society changes, this type of bias will have to be **manually compensated for** in algorithms

Biases in Training Data

- Bias may arise in algorithms if training data:
  - **Under- or overrepresents groups** in the target population
  - Was produced in conditions that were **overly restrictive** (initial Street Bump app)
  - Contains **protected characteristics** (race, national or ethnic origin, religion, sex, etc.)
  - Contains **proxies for protected characteristics** (geographic region)
    - If protected characteristic is predictively powerful, other data may infer it
- Bias may also arise if:
  - Available training data **does not match goal** of the system (COMPAS)
- Key measures to address this type of bias include **pre- and post-implementation validation studies, audits and peer review**

Biases from Practitioners

- Data scientists may introduce bias into algorithms when they make decisions about the **design, analysis and interpretation** of systems—for example:
  - Determining thresholds to map continuous variables to discrete outcomes / recommendations / activities
  - Setting error rates for false negatives and false positives
    - E.g., Philadelphia's Adult Probation and Parole Department's risk prediction algorithm set error rates to reflect a policy that it is "much more dangerous to release Darth Vader than it is to incarcerate Luke Skywalker"
- An important measure to address practitioner bias is to ensure a **diverse team from a range of backgrounds**—ethics, technology, law, privacy, community membership
  - Similar to the role of a research ethics board (REB)

Conclusion

- No “silver bullet” or automatic way to eliminate bias from algorithms
- Humans should always remain accountable and “in the loop”
- Like security, **risk-based approach** to algorithmic decision-making is important to consider:
  - The greater the impact of a decision on the rights or interests of an individual, the greater the requirements to ensure fairness, accountability and transparency

HOW TO CONTACT US

Information and Privacy Commissioner of Ontario

2 Bloor Street East, Suite 1400

Toronto, Ontario, Canada M4W 1A8

Phone: (416) 326-3333 / 1-800-387-0073

TDD/TTY: 416-325-7539

Web: www.ipc.on.ca

E-mail: info@ipc.on.ca

Media: media@ipc.on.ca / 416-326-3965