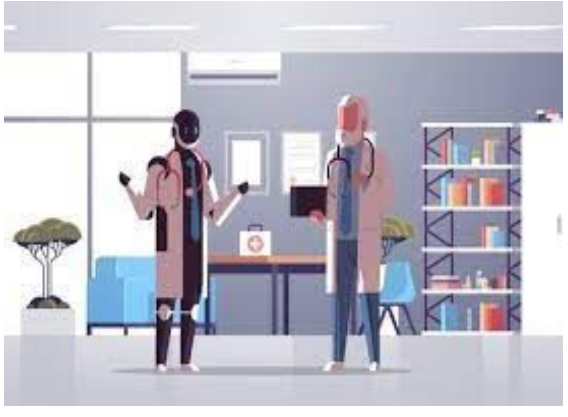
An introduction to AI, algorithms and ethics in the public service

CREDIT: GETTY

Mehrdad Safaei, Analyst, Data and AI Learning Products | Digital Academy
mehrdad.safaei@csps-efpc.gc.ca

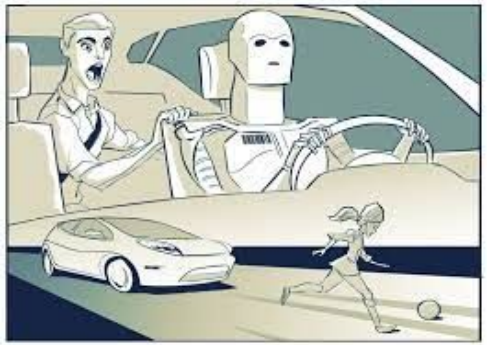
AI in the Government

- The Correctional Offender Management Profiling for Alternative Sanctions, or COMPAS, is a risk assessment algorithm used in a number of U.S. jurisdictions to assist in risk assessments for determining bail conditions (pre-trial) and sentence lengths (post-trial). COMPAS assigns defendants a score from 1 to 10 that predicts how likely they are to reoffend. The score is based on 137 variables that include age, sex, and criminal history (though race is not one of the questions).
- In Canada, Federal Department of Immigration, Refugees and Citizenship has used AI for several years to process immigration visa applications. This programmed algorithm approach applies codified immigration rules to applications to pre-process and triage them.
- In UK, Babylon Health is a digital-first health service provider that combines an artificial intelligence powered platform with virtual clinical operations for patients.



Ethics and Bias

1. Algorithmic bias describes systematic and repeatable errors in a computer system that generate unjust (or unwelcome) outcomes in areas such as hiring, security, elections, healthcare, and criminal justice. One of the earliest examples of algorithmic bias occurred at St. George's Hospital Medical School in the UK. From 1982 to 1986, 60 women and visible minorities were unfairly discriminated against and rejected because a new computer assessment system was introduced that was biased in favor of foreign-sounding-names based on historical trends in admissions. Thus, women and visible minorities without foreign-sounding names were rejected from the program.
2. Since COMPAS is apparently not in use in Canada, its constitutionality here has not been tested. However, section 11(e) of the Canadian Charter of Rights and Freedoms guarantees a right to bail unless reasons are given (and, presumably, “because the computer said so” would not be deemed a valid reason). Would such a requirement prohibit the use of COMPAS in Canada?
3. False Accusations and Arrests Based on Biassed Face Recognition
4. Privacy and Safety
5. Responsibility
6. Copyright

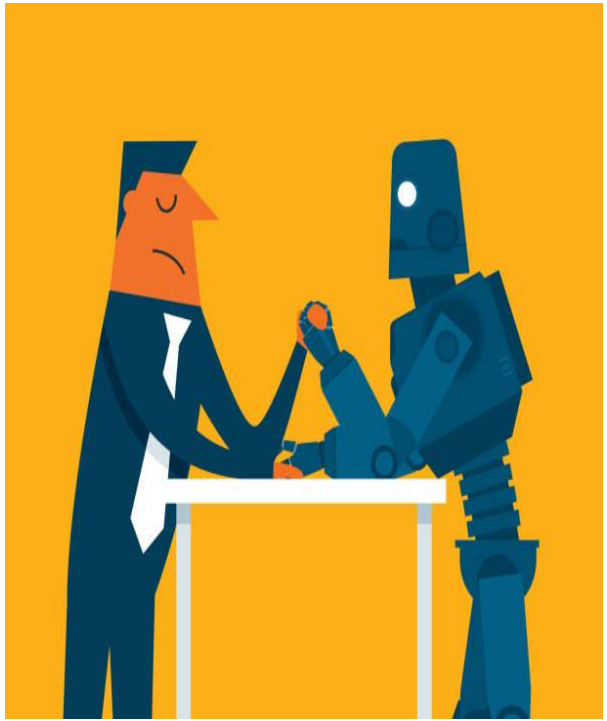


Reducing Risk

- The “right to an explanation” is a movement aimed at combating the black box problem of algorithmic decision making (where we have no way to understand how technology makes many basic decisions for us), requiring that if someone is affected by a decision made by an autonomous device or algorithm, that the responsible organization tell us how that decision was reached.
- The right to an explanation is included in the European Union’s General Data Protection Regulation (GDPR) at Articles 13–15 which provides rights to “meaningful information about the logic involved” in automated decisions, but is unlikely to present a complete remedy to algorithmic harms as the law is restrictive, unclear, or even paradoxical concerning when any explanation-related right can be triggered. Also, the legal conception of explanations as “meaningful information about the logic of processing” may not be provided by the kind of ML explanations that computer scientists have developed.
- An alternative approach is embodied in the Government of Canada’s Directive on Automated Decision-Making, describing how the government will use AI to guide decisions within several departments. Released on March 4 2019 to govern automated decision systems developed or procured after April 1, 2020. It includes a requirement for an “Algorithmic Impact Assessment” (AIA; modelled on environmental impact assessments, or EIAs) prior to the production of any automated decision system. The AIA is a tool that determines what kind of human intervention, peer review, monitoring, and contingency planning a proposed automated decision system or AI tool will require.

Explainability
Transparency
Interpretability

QUOTE

Kranzberg's first law of technology, "technology is neither good nor bad; nor is it neutral" (Kranzberg, 1986).

As shown in many examples, AI technology and machine learning models, including the data that they leverage, aren't necessarily more or less biased than humans that create or deploy them. Recognizing this is an important first step as an individual to helping ensure AI is as responsible as possible.