Bias in AI: Practical examples in Natural Language Processing (NLP) & Computer Vision (CV) for the CRA

Workshop Outline

General Information

Class Dates	**Orientation:** Tuesday, January 10, 2023 **Lectures:** Mondays from January 16 to February 20, 2023 **Tutorials:** Tuesdays from January 17 to February 14, 2023 **Capstone Presentation:** Monday, February 27, 2023
Class Times and Location	10:00 am to 12:00 pm EST Virtually via Zoom

Workshop Description

This workshop will build AI capacity among CRA staff both in technical skills and in responsible applications of AI. Bias in AI may exclude or disadvantage sections of the Canadian population, an undesirable outcome in a country as diverse as Canada. Providing education on how bias in AI can be identified and potentially reduced will empower CRA practitioners to mitigate against bias and strengthen their leadership in responsible AI. This workshop will be delivered in an interactive lecture/discussion format, with practical examples and case studies used to spark reflection and questions.

Level of Training and Prerequisites

Workshop assignments will involve coding in Python. Python is a high-level programming language, designed for easy readability, and supports the development of a variety of applications.

The workshop will expose learners to several open-source Python packages. Learners should have some familiarity with the following packages:

- `os` or `pathlib`: handles filesystem paths
- `numpy` and `pandas`: manipulate arrays and dataframes
- `matplotlib`: creates visualizations for exploratory data analysis or results

- `scikit-learn`, `TensorFlow` or `PyTorch`: popular machine learning frameworks

A pre-workshop assessment will gauge learners' understanding of the above Python packages and suggest resources for addressing any gaps in knowledge.

Assignments will be provided in the form of Jupyter notebooks (`.ipynb` files). Notebooks can be executed within a web browser using Google Colab or locally on a learner's personal computer. Learners should be familiar with navigating the notebook environment, writing and executing Python code, and styling text using Markdown. A brief orientation will be given at the beginning of the workshop.

Our teaching team recommends the use of Google Colab during the workshop since it removes the need to install any programs on the learner's computer, provides a uniform environment for everyone to work from, and provides free access to GPUs for executing more computationally-intensive code.

Note: Learners won't be expected to write NLP or CV code from scratch. However, they will be expected to understand the code and modify parameters for tuning the models.

Learning Outcomes

- Obtain hands-on experience working with ML models for both NLP and image-to-text, and tuning model parameters
- Evaluate AI technologies and discuss the unanticipated or unwanted impacts of biases through several case studies
- Learn about strategies for identifying and measuring bias, with a focus on NLP
- Understand both technical and non-technical strategies to mitigate bias
- Discuss the multidisciplinary nature of AI adoption and policies to avoid/mitigate bias
- Be empowered to apply concepts and techniques into their day-to-day work

Workshop Schedule

Workshop Welcome & Orientation Jan. 10 – 15	• Post on the "Let's Get to Know You!" discussion board • Complete entrance survey • Complete Python refresher assessment • **Due:** before the first class (10:00 am EST on Jan. 16)

Week 1: Introduction to Bias in AI
Jan. 16 – 22
 	
• What is AI and why does it matter for the CRA?
• What are the key challenges in AI?
• What is bias?
• What are some commonly seen examples of bias in society? In AI?
• How is bias introduced into AI? (for example, data, algorithmic bias)
• How can we measure bias?
• **Discussion:** Is bias always bad?
	• **Tutorial:** Discussion on bias in AI
	• **Tutorial Discussion:** Consider how bias affects CRA and the federal government more broadly
	  ○ How might bias be demonstrated in different governmental datasets?
	  ○ What issues are you most concerned about?
	• **Case study** on decision support tools in government
	• Introduce capstone project outline and expectations using a business lens
	• Participate in class discussion and post on Discussion board before next class (10:00 am EST on Jan. 23)

Week 2: Introduction to Natural Language Processing (NLP) Jan. 23 – 29 • What is NLP and how can it be applied? • What are the major challenges with NLP? • Data is unstructured and requires pre-processing; what needs to be done before data are input into an NLP model? • How can we measure bias in NLP? • What are the popular language models for NLP?	• **Tutorial:** BERT model and identifying bias in NLP models • **Assignment 1:** Work with a pre-trained BERT model and visualize different types of bias it learned • **Due:** before next class (10:00 am EST on Jan. 30)
Week 3: Introduction to Computer Vision (CV) Jan. 30 – Feb. 5 • What is CV and how is it applied? • What are the major challenges with CV?	• **Tutorial:** Text detection and recognition in images • **Assignment 2:** Learn how to use a simple text recognition model (PDF to text) and create your own model

• How is data organized for CV? • What are the popular models for CV?	• **Due:** before next class (10:00 am EST on Feb. 6)
Week 4: Mitigating Bias in AI Feb. 6 – Feb. 12 • Data collection/organization methods • Governance, privacy and ethics • Introduce strategies for identifying and mitigating bias • **Discussion:** Consider recent policies put in place by government and other institutions to mitigate bias in AI	• **Guest speaker** on detecting and mitigating bias • **Assignment 3:** Work with practical strategies for mitigating bias in an NLP model - recycle example from Assignment 1 • **Capstone:** Submit brief project proposal for interim feedback • **Due:** before next class (10:00 am EST on Feb. 13)
Week 5: Mitigating Bias in AI cont. & Summary Feb. 13 – 19 • Continue strategies for mitigating bias • Summarize workshop content	• **Guest speaker** on detecting and mitigating bias
Week 6: Capstone Projects Due Feb. 20	**Submit your capstone project by 11:59 pm EST on Feb. 20)**
Week 7: Capstone Presentation Feb. 27	**Oral presentation of capstone project**