Assignment 3 – Bias: Technical and Policy Considerations

Assignment Instructions

Introduction

In this assignment, you will consider how bias affects your organization and industry, and you will consider technical and policy mitigation strategies to reduce bias.

Part 1: Bias in Business - Technical Considerations [50 points] [Min 500 words, Max 2000 words]

Analyze an AI or data driven application in your organization, your own product, a product you wish to develop, or a product you have worked on in the past. If you are unable to share an example from your business, you may refer to the NLP or Computer Vision assignments as though they were deployed in industry. Find the potential areas where bias is introduced. In your writeup, answer the following questions:

- What is the source of bias (in each stage of the cycle)
- What type of biases are there?
- Are the biases harmful (socially)? If yes, what type of harm?
- Is there any network effect?
- Can you think of the ways to reduce the bias?
- What would be the cost of bias reduction?
- What immediate steps would you take to reduce bias?

You will need to provide a brief paragraph about your product and the context for your analysis.

Part 2: Bias in Business - Policy Considerations [50 points] [Min 500 words, Max 2000 words]

AI adoption is a multi-disciplinary endeavour. AI practitioners must often communicate their technical approaches and considerations to non-technical audiences to advance AI adoption in an organization and among stakeholders. Businesses and Governments must design frameworks to make the use of AI safe and fair and these policies must translate into technical environments and must guide technical implementations such as the two assignments you have submitted so far. Given the highly technical nature of the technologies,

technical practitioners such as yourself are increasingly required to advise business leaders and policy makers on policies frameworks and guidelines relating to AI.

Consider the following policies/frameworks and write a draft policy/framework to advise your organization on how to avoid and mitigate bias in your AI or data product offerings. You will need to provide a brief paragraph about your product or industry and the context for your guidelines.

Here are some policy/framework examples:

- https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/ – "We are pleased to announce AI Fairness 360 (AIF360), a comprehensive open-source toolkit of metrics to check for unwanted bias in datasets and machine learning models, and state-of-the-art algorithms to mitigate such bias. We invite you to use it and contribute to it to help engender trust in AI and make the world more equitable for all."

- https://ai.google/responsibilities/responsible-ai-practices/?category=fairness – "We believe that these technologies will promote innovation and further our mission to organize the world's information and make it universally accessible and useful. We recognize that these same technologies also raise important challenges that we need to address clearly, thoughtfully, and affirmatively. These principles set out our commitment to develop technology responsibly and establish specific application areas we will not pursue."

- https://open.canada.ca/aia-eia-js/?lang=en – "Algorithmic Impact Assessment (AIA), is a questionnaire designed to help you assess and mitigate the risks associated with deploying an automated decision system. The AIA also helps identify the impact level of your automated decision system under the Directive on Automated Decision-Making."

What To Submit

You must submit one PDF file answering all questions from both Part 1 and Part 2. The file should be saved as lastname_assignment2.pdf. Submission should be made through D2L.

This report is preparatory work for your Capstone Project so you may conduct model implementations to support your analysis, however, this is not a requirement. Include any relevant graphics in your PDF. Your writeup must be typeset and you must cite any sources you use.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ..., Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[2] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

[3] Ba, Jimmy Lei, Jamie Ryan Kiros, and Geoffrey E. Hinton. "Layer normalization." arXiv preprint arXiv:1607.06450 (2016).

[4] https://blog.einstein.ai/the-wikitext-long-term-dependency-language-modeling-dataset/

[5] Liu, H., Wang, W., Wang, Y., Liu, H., Liu, Z., Tang, J. (2020). Mitigating Gender Bias for Neural Dialogue Generation with Adversarial Learning. EMNLP.

[6] Maze, Brianna, et al. "Iarpa Janus benchmark-c: Face dataset and protocol." 2018 International Conference on Biometrics (ICB). IEEE, 2018.

[7] Arriaga, Octavio, Matias Valdenegro-Toro, and Paul Pl¨oger. "Real-time convolutional neural networks for emotion and gender classification." arXiv preprint arXiv:1710.07557 (2017).

[8] http://chalearnlap.cvc.uab.es/challenge/38/description/

Assignment Developed by Shingai Manjengwa and Sayyed Nezhadi

Instructor: Sayyed Nezhadi | Course TAs: Anastasia Razdaibiedina |Rishav Raj Agarwal

| Course Director: Shingai Manjengwa (@Tjido)