Capstone Project Plan

Last updated: November 19, 2023

Instructions: Each capstone team can use this template to capture and summarize information about the project. This can be shared with the sponsor and mentor. When submitting the plan during the course, a PDF file is preferable.

Stakeholder Names and Roles

Stakeholder	Role
Becky Desrosiers	Team member
Abner Casillas-Colon	Team member
Naomi Ohashi	Team member
George Shoriz	Team member
Phillip Waggoner	Mentor
Emanuel Moss	Sponsor
Elizabeth Watkins	Sponsor
Dawn Nafus	Sponsor

Project Title: Bias Evaluation in Open Model Platform

Abstract

This project seeks to evaluate one model in Intel Labs' Open Model Zoo for potential bias against protected characteristic(s). The team will start by researching bias metrics, identifying useful datasets, and choosing a model that will be feasible to evaluate, with appropriate metrics. Finally, the model's inference on the datasets will be evaluated using the chosen bias metrics.

Outline of the Project

Intel's Open Model Zoo (OMZ) is an offering that allows regular people to use pretrained AI models for their own purposes. This project is important because AI models can easily have bias inherent to their training, which can impact their performance and their impact on society, depending on for what purpose they are leveraged. Since the OMZ is publicly available, the potential for applications is

expansive and, in turn, so are the potential for consequences from biased training. Stakeholders could include anyone who employs the model, or anyone who could be affected by myriad applications of the model.

The scope of this project is one model, and finding a metric or set of metrics that can quantify the bias in the model's training. We assume that we start with a possibly biased, pretrained model. A stretch goal of the project will be to create a pipeline that will facilitate future bias detection in other models.

Success Criteria

SC1	Summary of existing bias metrics for AI models
SC2	Identified datasets with ground truth that can be used with the chosen model
SC3	One model summarized in terms of chosen bias metrics (Final report and presentation)
SC4	(Stretch goal) Initial development of testing pipeline to follow the lead of this project

Assumptions and Limitations

For any project, there may be assumptions [A] *and limitations* [L] *on the data and the modeling approach. These can be documented here. Example:* [L] *Ideally, the dataset would include variable X, but we did not have access to this data, which was a limitation.*

Identifier	Description
	This section to be more fleshed out as we find datasets and choose a model

Potential Background Literature and Resources

- "The four-fifths rule is not disparate impact: a woeful tale of epistemic trespassing in algorithmic fairness" https://arxiv.org/abs/2202.09519
- "Measuring Model Biases in the Absence of Ground Truth" https://dl.acm.org/doi/pdf/10.1145/3461702.3462557 and references
- Project GitHub repository https://github.com/oatmeelsquares/BiasOMZ
- Google Dataset Search https://datasetsearch.research.google.com/

Brief Outline of the Data

TBD

Brief Plan of How the Data will be Modeled and Processed

We will be using data that is publicly available online. There will be no need to store it long-term. Processing will depend on what datasets we find.

Brief Plan of Modeling Approaches

We will use the OMZ model as intended. Which model to use is TBD. We may use metrics such as precision and recall to compare the differences in performance between different demographics. However, the final metrics we will use is TBD.

Potential Concerns [C] and Blockers [B]

Identifier	Description
C	Keep in mind the nuance of metrics, that rarely can things be encompassed in a single number
C	Potential block if we cannot find data labelled with demographic-related ground truth. May need to crowdsource labeling or use a model that also predicts the demographic.

Final deliverables

A written report with:

- literature review including what techniques currently exist for detecting bias, their strengths and limitations
- Description of the data & how we found it
- Description of the model and what metric(s) and/or libraries we used to evaluate it, based on the lit review
- Results of evaluation
- Conclusion: does the model show bias or not?
- Potential: description of an established pipeline OR recommendations for evaluating future models

A presentation with:

- What we found in our lit review and why we chose the metrics we did
- Description of the model we chose and why we chose it
- Description of the process of getting the data, accessing the model, and evaluating it
- What we found - performance on bias metrics
- Conclusion: biased or not?
- Potential: description of an established pipeline OR recommendations for evaluating future models

Project roadmap

Week 1 (Jan 17-23):

- Determine final deliverable
- Set out project roadmap (this)
- Email sponsors with project plan and roadmap

Weeks 2-4 (Jan 24 - Feb 6):

- Dig into researching bias in AI and possible metrics
- Add notes to the issue

Week 5: (Feb 7-13):

- Tentatively decide on what metrics and model to use
- Find datasets to test on

Week 6: (Feb 14-20)

- Finalize decision on what metrics, model, and dataset(s) to use
- DETAILED documentation of decision and reasoning
- Begin writeup of the previous

Week 7: (Feb 21-27)

- Finalize writeup of Report 1

Feb 27: Report 1 DUE

Week 8: (Feb 28 - Mar 6)

- Get data into (jupyter) environment

Week 9-11: (Mar 7-27)

- Spring break
- Implement tests, evaluate model
- DETAILED documentation of findings

Week 12: (Mar 28 - Apr 3)

- Flex week - extra time for implementing tests/evaluating model if things go wrong, or starting to establish the bias eval pipeline
- Begin writeup of report 2: what we have accomplished so far and plan for pipeline
- May begin work on the pipeline in this week

Week 13-14: (Apr 4-16)

- Write/finalize report 2
- Start on pipeline in earnest
- DETAILED documentation of pipeline

Apr 17: Report 2 DUE

Week 15: (Apr 17-23)

- Finalize pipeline
- DETAILED documentation of pipeline
- Start final report writeup

Week 16: (Apr 24-30)

- No more technical work (it is what it is)
- Final report
- Final presentation

May 1: Final Deliverables DUE