A reflection on bias, fairness, and environmental impact during my two projects at the California Academy of Sciences

Maricela Abarca

Introduction

My practicum work has involved two main projects for the Cal Academy's Scientific Computing department. The first was completed over the first three months and entailed analyzing server usage logs to understand trends in research compute needs. The goal was to purchase new machines that best increased research capacity by assessing how staff and students were using the current configuration. The utilitarian lens, fairness for different users, and environmental impact due to energy consumption are some of the ethical considerations I've reflected on below for this project. As that work wrapped up, I began on my second project: building a tracking system for institutional publications to replace the current manual method of recording publications. My task was to automate this process and ensure that publications were collected in a standardized, organized, and reproducible way. Although this project was presented to me through a utilitarian lens, I reflected on it through the lens of justice/fairness. I felt that bias and fairness were important to consider since there was potential for disparate impact on some groups, such as early career researchers.

Ethical Considerations and Decisions

The server analysis project seemed fairly benign at first in terms of potential negative ethical consequences. From preliminary analyses, it appeared that any new server purchase within our budget would help meet the needs of users by adding more memory and compute power. We approached the project with this utilitarian motivation, but also stopped to consider the downstream fairness impacts of any incorrect conclusions we could draw. For example, we asked ourselves how underestimates of usage would impact outlier users who need access to larger amounts of compute power to complete their research. We wanted to make sure we were capturing overall trends to satisfy the most people, while still considering these high outlier users. In order to achieve this, we assigned greater weights to RAM and core number in our optimization function, but then went back and checked the returned configurations to make sure they could handle all compute usage cases appropriately. Looking back on our optimization parameters, we could have included efficiency in the equation in order to consider the energy consumption and environmental impact of the new purchase. These machines pale in comparison to the scale and energy consumption of the AI compute resources described by Melissa Heikkila in the article *Making an image with generative AI uses as much energy as charging your phone,* but that doesn't necessarily excuse overlooking the environmental impact

of our server configuration decision. In the future as the research compute cluster grows, efficiency can and should be considered for future purchases, especially since the Cal Academy holds sustainability as a core value. This approach can help the Scientific Computing department not only support staff research but also incorporate sustainability into its everyday functions.

In contrast to the server usage project where I didn't immediately think of any ethical red flags, I recognized the potential negative impacts of the publication tracking system from the outset. The project was described through the utilitarian perspective, with an emphasis on the overall good that would come from it. It certainly felt that anything would be an improvement over the current manual and error prone way of collecting publication metadata, but issues around underrepresentation, which could negatively impact people's professional evaluations, were still apparent to me. I don't think I had the language yet to articulate my concerns, but the ideas circulating in my mind aligned with what we later read and learned about with respect to the justice/fairness ethical lens and algorithmic bias and fairness.

For example, in the *21 Definitions of Fairness* video, Arvind Narayanan mentions that different metrics matter to different stakeholders. The stakeholders for my project are researchers and the people who evaluate them. Although the best possible publication coverage is likely desired by both groups, evaluators may diverge from researchers on what other metrics are important for assessment (e.g. mentorship hours, public engagement, etc.). I admittedly struggled with being tasked to only focus on publication metrics, since that feels like a very one-dimensional metric that also doesn't reflect fairly on different groups. However, we had to start somewhere in improving the current state of publication tracking, so I tried to make the tool as comprehensive as possible and communicate my results along the way. As I developed the system, I shared spreadsheets of the publication results I was retrieving with both stakeholder groups. I was motivated by wanting to be transparent and accurate. In retrospect I was incorporating some of the "Justice/Fairness-Related Questions" from our first reading. These included consulting stakeholders so that their feedback could help guide the project, considering the benefits versus burdens of the project, and weighing the risks of harm versus disproportionate benefit. I've also done my best to convey any potential shortfalls of this project in collecting all possible publication metadata, and to make clear that publications alone should not represent an individual's overall impact and value to the institution. I've suggested other future data aggregation projects to capture metrics that would hopefully give a more well rounded perspective on staff impact as well.

Although we learned about bias and fairness in machine learning contexts and I wasn't tasked with implementing any machine learning models, I think the concerns around disparate impact still apply to this project. When we discussed COMPAS in class, it became apparent that we should pay close attention to effects on groups that have the most to lose when considering applications of new technology. In my case, this would be early career researchers and students since these groups rely on publication metrics (for better or worse) to advance their careers, but their academic accomplishments often get overlooked. I wanted to make sure that I wasn't trading the disparate impact of manual tracking with another version of disparate impact in the

form of automated metadata collection. I internally prioritized making sure that I was getting the best coverage possible for these sensitive groups and spent the extra time to realize this goal.

Conclusion

This class has given me the language to go back and reflect on various ethical perspectives of my practicum work. For my server usage analysis project, I was able to identify an overlooked ethical consideration and become more mindful of the environmental impacts of my work. In revisiting the publication tracking system, I was able to better articulate my initial reservations about the project from the perspective of ethical lenses and definitions of bias and fairness. I don't think that any of the ethical considerations I've described for my projects are particularly egregious compared to some of the examples we discussed in class and that were in our readings, but that doesn't mean the Cal Academy can't be proactive in addressing ethical concerns of scientific computing work going forward. One emerging theme of our readings that I noticed is that it's best to be proactive with risk instead of reacting when things go wrong, so it couldn't hurt to consider environmental impact with server purchases or disparate impact of evaluation tools at the outset of these endeavors.

I think my own career will be shaped by data and AI ethics in a variety of yet-to-be-determined ways. Although I'm not sure where I will land on the career spectrum of data science roles that contribute to the implementation of models and AI systems, I do feel that I have the proper ethical scaffolding to help me think critically about data, both in and out of the workplace. I also feel that I have some good anchor points to discuss these topics in the varying landscapes of organizational ethics. Since I'm interested in climate modeling and could see my career intersecting with this domain, I would be curious to see how organizations think about data ethics when it comes to fairness and bias in applying AI to climate solutions. But before I even get a job, I'm sure my career is being shaped already by AI ethics in the form of automated resume scoring and other candidate screening tools.