Annotation and Benchmarking on Understanding and Transparency of Machine learning Lifecycles (ABOUT ML)

Chapter 3: Current Challenges of Implementing Documentation

This section is where PAI invites comments, anecdotes, case studies, broader stories from implementing documentation efforts, and results from any solutions (effective or ineffective) attempting to address these challenges. Please also share feedback on whether your organization has encountered these challenges or new ones, or if these challenges do not exist in your work.

When attempting to implement the recommended documentation guidelines, a number of common challenges arise. The following is an overview of currently identified challenges. The eventual goal of this chapter is to help practitioners foresee challenges in their own settings and provide solution options for addressing them.

Benchmark Customization: For many documentation criteria, it is very difficult to identify appropriate and demographically representative benchmarks to completely evaluate a model system. Often, separate custom benchmarks specific to the ML system's context of use need to be developed for the documentation requirements and the evaluation of deployment. However, this can be costly and time intensive, so there are real organizational tradeoffs to navigate between benchmark quality, speed to deployment, and quality of evaluation and documentation.

Metric Selection: There exist numerous academic and industry metrics to measure the performance of an ML system and many fairness evaluation metrics and definitions. It is currently difficult to say without further understanding of each unique situation and context which metrics are appropriate or best. Thus, metric selection is additional work that teams need to budget time for. In "Machine Learning that Matters,"[1] Kiri Wagstaff suggests the high level guidance of defining metrics according to the intended outcome rather than evaluating model performance on arbitrary test sets with typical ML performance metrics, which is a reasonable starting point for most projects. For instance, if an ML system is meant to optimize for revenue, then measure revenue outcomes from the ML system directly, or make use of a proxy like advertisement impressions, rather than rely on the Area Under The Curve (AUC) of an isolated model.

Privacy - Soliciting and mining the demographic metadata used in evaluating whether a model system is performing fairly across intersectional subgroups can also expose identifying information about users and image subjects. To mitigate this risk, it is important to store the data in a way that respects privacy and does not compromise individual privacy in exchange for ML system-level transparency.

Intra- and Inter-Organizational Cooperation - It is very difficult to set up the novel organizational processes, get buy in, and secure monetary and HR resources required to

[1] Wagstaff, K. (2012). Machine learning that matters. arXiv preprint arXiv:1206.4656. https://arxiv.org/abs/1206.4656

effectively adopt documentation for transparency as an organizational norm without cooperation and alignment across multiple levels of internal and possibly external teams. Getting alignment around prioritizing the principle of transparency is the critical first step to implementing any documentation practices. Given the early stage of implementation for documentation practices, organizations may also need to look to outside expertise to aid in the process of designing the right processes, templates, and practices. ABOUT ML hopes to offer a starting resource in this.

Compromising Intellectual Property - One commonly feared risk of documentation is losing trade secrets and intellectual property due to disclosing too much information. However, allowing protecting intellectual property or preserving "trade secrets" to serve as a blanket excuse for omitting information in documentation opens the door for companies to hide crucial information that should be revealed in the public interest. One goal of ABOUT ML guidelines is to indicate areas all companies should be willing to share. In each documentation recommendation section in Chapter 2.4, we discuss more specific pros and cons of disclosing that specific information about an ML system to better inform this tradeoff.

System Security Vulnerability - Another fear for documentation is revealing attack surfaces in an ML system by providing too much insight into how it was built. This is a fine line to walk because on the one hand, it is an organization's responsibility to build robust security measures into their ML systems and documenting these may spread more knowledge for other organizations attempting to secure their own systems. On the other hand, people fear that knowledge being misused for hacking rather than for shoring up collective defenses. It is worth a detailed discussion as an industry to better discern between low and high risk types of information disclosure. Additional security risks include documentation that reveals potential blindspots of the model system such that nefarious actors could game or hack the system. The first step to finding solutions for these risks is by naming them, and the next step involves investing in research and further understanding towards best practices.