Ethics Report

In this report, we will reflect upon different ethical aspects that concern our extraction and aggregation of data from OpenReview.net. Since the data we retrieved is not manipulated or used in any Artificial Intelligence or Machine Learning Technology by our group, only the implications of the raw data are considered. In addition, the combination of our data with data acquired from different sources should also be considered to point out possible abuse scenarios. Group 2B, who use our data in their follow-up project, will have a more detailed analysis on the ethical considerations regarding the data inferred from our dataset.

The intended use of our dataset is the comparison of paper revisions and matching the changes to comments and reviews on the paper. For this purpose, the data needs to include paper submissions, revisions, comments and timestamps of revisions and comments. Our data goes beyond this information, for the sake of completeness.

The following sections are considered relevant for a closer analysis:

1.) Privacy

2.) Bias

3.) Further Abuse Scenarios

1.) Privacy

Author names are kept in the dataset for both comments and papers as far as they are not already anonymized by OpenReview to begin with (blind submission, anonymous review, etc.). They are not anonymized in any way by us. Although the names are not required in the follow-up task for group 2B, they were included for completeness and possible future tasks. For example, with the Athena Chat Bot it is necessary to keep the names so questions can be asked about a specific person.

The acquired information in our dataset is available on OpenReview.net and can be viewed without special access requirements. Everything is also available with their official API which can be used without limits by everyone. Information on venues, papers and authors can usually also be gained through different means. For example, venues usually have a separate webpage to inform attendants or the public. Authors who publish their papers consent to having their name associated with their work. The papers uploaded to OpenReview are intentionally regarded as “openly published” and we assume the authors are aware when uploading their material.

Also, as we are dealing with academic papers, authors usually want their names to be used in combination with their (accepted) submissions and it can be seen as disrespectful to leave out the authors. Papers are usually cited by author names and not titles. The case of rejected papers is slightly different but it is possible that the author might want to attempt to publish an adjusted paper at a later date or to a different venue.

Only anonymizing rejected papers can introduce a bias in the data. In addition, there are papers where the acceptance status is unknown or where the label might be assigned incorrectly by us. This might then anonymize an accepted paper by mistake which should be avoided. The same considerations apply to withdrawn papers.

Comments can usually be posted anonymously and reviews are usually anonymous as well. Comments by the authors are usually marked as such so they are not anonymous but then the above applies regarding anonymisation by us.

However, there are a few data points that could potentially be used to acquire further information that was not intended to be disclosed. In the process of writing and adjusting a submission, the author may edit submissions and reply to comments or reviews. Revisions of submissions and especially comments are stored by OpenReview.net and play an important role in our dataset. Comparing these revisions could be exploited to gain insight on the author's writing style (use of vocabulary, grammar, other linguistic features). As an example, the writing style of author X is analysed by a machine learning algorithm. A new paper is published anonymously for reviewing. It is theoretically possible to match the style of the anonymous paper to an author, thus violating the anonymity and possibly sabotaging the venue's double blind review process. Doing this would require a big amount of data and one would have to be able to discern between groups of authors and individual authors. Although this seems like a highly unlikely scenario and not much benefit is to be expected by such a procedure, it is still important to mention, because this is only made possible by keeping the names in the dataset.

Another possible breach of privacy can happen when having a closer look at comments and their revisions. Old versions of comments can be viewed and are directly attached to the comment in the latest version. Comparing these comments could be used to learn personal traits or habits, even if they are just typographical error or style changes. While this raises some ethical concerns, keeping revisions visible is also the policy of OpenReview. Therefore, the information is included in our dataset.

2.) Bias

The ACM Code of Ethics[1] states that a computer professional should "be fair and take action not to discriminate". Again, since our data is true to the original data, there is supposedly no bias created by crawling.

The only bias possible is due to the data distribution itself: old conferences and workshops have on average fewer submissions, e.g. ICLR and NeurIPS have far more submissions than other venues and thus their themes dominate other topics and so on. Should our data be used in ML or AI tasks, it is possible that a bias is created to a certain degree due to the inherent bias of OpenReview.

Different bias can also appear from use of the data, for example using the author names. There is no personal information on authors included in the data, but our data can be linked to resources about the authors. Names can be used to infer information about gender or ethnicity (although this is very speculative and should be reconsidered). If the email address is given, the university and possibly country of residence could be used in a technology. As a fictional example we look at an "automatic review score" algorithm, which from the given data learned that papers from University X or even Country Y are more likely to be rejected at conferences. Even though the facts are a true representation of the data points, it shows a strong bias. Such an algorithm can be very harmful. This is very similar to the prominent example of having a mortgage denied because of racial bias in the scoring algorithm[2].

[1] https://www.acm.org/code-of-ethics

[2] BARTLETT, Robert, et al. *Consumer-lending discrimination in the FinTech era*. National Bureau of Economic Research, 2019.

3

3.) Further Abuse Scenarios

A few general considerations concern the quality of the data. This is due to the fact that OpenReview.net is not used by *every* venue and mainly focusses on venues in the research field of machine learning. It is important to state that only a few selected venues, and within those only some years make use of the reviewing platform. The extent to which OpenReview is used also differs greatly between venues and years. Some more recent or upcoming venues have a good amount of comments and reviews, while others have very few or none.

In general, this raises questions about the representativeness of our data. Should the data be used to reason about a reviewing and acceptance process of a venue, it is important to keep in mind that not every venue has published their process on OpenReview.

Overall, the domain of the dataset is very specific and there is an inequality of content amount present. Therefore, the usefulness strongly depends on the application. Any future users of our dataset should be aware of these concerns and also mention it when publishing results based on the data.

3