General response. We thank the reviewers for their helpful comments and remarks. We are glad that you appreciated our efforts for the scientific clarity and the experimental description. We are pleased that you acknowledge the mix between theory and practice in our work.

We note that reviewers are interested in extensions of the current paper to a wider range of tasks and corpora: Reviewer 1 states that *'Sentiment analysis is not the most compelling task; it would be good to show the generality of this method on other tasks'*. Reviewer 3 suggests to assess *'if the data augmentation scheme is actually helpful for a range of tasks'* and Reviewer 4 mentions to work with *'a larger set of corpora'*.
↪ We agree. In this paper, we introduce **LHTR** and **GENELIEX**. We worked on two datasets, *Amazon* with 231k reviews and *Yelp* with 1450k reviews as stated on line 261, and further detailed in Appendix B.6. The focus is here on a method for learning a heavy-tailed representation (including a data-augmentation procedure) and its impact on a standard text classification task, namely sentiment analysis and on data augmentation. Table 2 shows that when the training set is augmented with **GENELIEX**, the classification performance increases (higher F1 score) for both *Amazon* and *Yelp* datasets (both medium and large size). Also, when compared with other methods, we achieve better performance. Future work will be the opportunity to address a wider range of tasks as it will also be the opportunity to work on a larger set of corpora, extending the work presented in this article.

Reviewer 1. Thank you for spotting the typos on lines 11, 58, 88 and Table 1. We will update the paper accordingly.

● *'Question: are the F1 scores here (Table 2) stratified at all by "extreme" vs not?'*
↪ A fasttext classifier is trained on the augmented training set (with various methods including **GENELIEX**). Table 2 reports fasttext F1-score computed on the whole test set with no special treatment made for extreme samples. Thus no stratification sampling is required.

● *'The proposed method regularizes fixed BERT embeddings to be heavy tailed, however most SOTA methods fine-tune the entire BERT model (instead of just an MLP on top). Though it seems like the proposed losses can still regularize these representations (affecting BERT + ϕ now instead of just ϕ), it's unclear if it will be as effective. It would be good to note whether or not fixing BERT is just for computational efficiency or not.'*
↪ BERT embedding was fixed for both computational efficiency and for evaluating the improvement solely resulting from ϕ in our experiments. We plan on fine-tuning the entire BERT + ϕ model during the training phase in future work.

Reviewer 2. Upon acceptance, the additional ninth page will be the opportunity to include the figures describing **LHTR** and **GENELIEX** content currently in the supplementary material.

Reviewer 3. Thank you for spotting the typos in Table 1. We will update the paper.

● *'There are no qualitative examples of outputs—this would especially be useful to see in this setting, for different words at the end of the distribution.'*
↪ Please refer to Table 8 in Appendix B.7.2. for output examples generated by **GENELIEX**.

Reviewer 4. We agree that text augmentation would particularly benefit from our embedding's scale invariance which is foundationally stronger than known token perturbation methods.

● *'How does the alpha value impact performance would be good to see here.'*
↪ The alpha value corresponds to the tail index of Z's heavy-tailed distribution. As the tail index increases, the tail gets lighter. It results that the greater alpha is the less likely it is for extremes to occur. Although we highlight that the approach is generic, in our experiments, the selected distribution is a multivariate Logistic distribution (see l. 220 and Appendix B.4). Other heavy-tailed distributions (with different tail indexes) may be selected.