**Deep Learning Mini-Project Datasets**

*English (Base Language)*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Link | Size | Remarks |
| Learning From the Worst (Dynamically generated hate speech dataset) | [Publication](https://aclanthology.org/2021.acl-long.132/)  [Dataset](https://github.com/bvidgen/Dynamically-Generated-Hate-Speech-Dataset) | 41’225 | Multi-category hate speech detection |

*Arabic (Training Data for Finetuning)*

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Link | Size | Remarks |
| Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech | [Publication](https://arxiv.org/abs/2201.06723)  [Dataset](https://sites.google.com/view/arabichate2022) | ~13’000 | This dataset is taken as a reference to embed cultural sensitivity as suggested in the paper **“Hate Speech Classifiers are Culturally Insensitive”** |