

Deep Learning Mini-Project Datasets

English (Base Language)

Name	Link	Size	Remarks
Learning From the Worst (Dynamically generated hate speech dataset)	Publication 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 Dataset	~13'000	This dataset is taken as a reference to embed cultural sensitivity as suggested in the paper " Hate Speech Classifiers are Culturally Insensitive "