

Label Counts by Language — Llama

Label

Counter-Speech	160	222	104	109
Neutral	34	29	29	25
Refusal	85	48	167	166
Hate Speech	21	1	0	0
	Albanian	Arabic	Chinese	English

Language