

Label Counts by Language — ChatGPT

Label	Counter-Speech	Neutral	Refusal	Hate Speech
	277	22	1	0
	295	5	0	0
	205	84	11	0
Language				
Albanian				
Arabic				
Chinese				
English				