

Label Counts by Language — Claude

Label	Counter-Speech	Neutral	Refusal	Hate Speech
	270	17	13	0
	270	6	24	0
	274	22	4	0
Language				
Albanian				
Arabic				
Chinese				
English				