

Label Counts by Language — DeepSeek

Label	Counter-Speech	Neutral	Refusal	Hate Speech
	291	9	0	0
	298	2	0	0
	288	12	0	0
Language	283	3	14	0