

Layer	Lang	Feature ID	Top Tokens
layers.0.mlp	English	93066	.'], !. , torino, erotik, ."</, " , porno, zásob, guards , -REAL
layers.0.mlp	English	128132	☹☹ , ☹ , !. , ?! , '. , ?" , " , , ... , !
layers.2.mlp	English	14207	http, Study, Henry, study , äh, háp, stone, https, Stone, http
layers.4.mlp	English	34138	耳, Accessible, üt, melt, ULLET, Weather , ialis, -gl, unction, Empty
layers.8.mlp	English	123973	ian, éné, jan , şa, Major , n, Robertson, duck , aira, ategorie
layers.15.mlp	English	5889	., In, , The, J, S, D, For, De, To
layers.15.mlp	English	67343	Manitoba, Tahoe, Elon, Newfoundland, Maui, Yosemite, Alibaba, Belg, Bolivia, Winnipeg