

Layer	Lang	Feature ID	Top Tokens
layers.0.mlp	Hindi	17276	aca, □, theless, گ, osit, avier, irable, emas, řen, 毒
layers.0.mlp	Hindi	26755	ceph, пю, enti, period, raf, rah, utt, ral, ersh, rosso
layers.0.mlp	Hindi	33080	bore, Cust, cek, nel, pedo, Cru, feu, ro, suite, enance
layers.0.mlp	Hindi	53040	,right, dle, plank, Richt, RIGHT, hton, sinh, triang, ões, wipe
layers.0.mlp	Hindi	55180	",[, fours, icone, .tif, agrams, saja, @a, Lawn, -wife, ubic
layers.0.mlp	Hindi	56263	immortal, grand, oux, isé, ideal, cratch, cout, VD, odynam, laid
layers.0.mlp	Hindi	76485	isphere, aclass, ailer, ahkan, onaut, ोत, ories, each, tables, や
layers.0.mlp	Hindi	101208	通り, imson, keh, Boca, 려, ksi, 平成, инку, ldre, urdu
layers.0.mlp	Hindi	122265	adesh, FromArray, alleng, aland, 站在, spread, slow, cido, Ama, thấp
layers.0.mlp	Hindi	128987	snapping, φι, 正, wen, Vernon, uffling, Cro, sparks, Cro, useless
layers.1.mlp	Hindi	9891	날, ward, Kern, ând, Tall, □, EdgeInsets, Lowe, abouts, Hull
layers.1.mlp	Hindi	49724	ocrat, eling, INGER, енд, asz, ucer, ground, ople, uce, mland
layers.1.mlp	Hindi	56943	hare, trap, calor, яч, trapped, everyone, ache, àn, andes, ho
layers.1.mlp	Hindi	65640	aves, Tib, educt, \$, urance, atab, plied, levance, 味, Beauty
layers.1.mlp	Hindi	106597	□, snatch, ères, rollo, umd, Alien, Win, lys, wo, bare
layers.2.mlp	Hindi	49394	arker, pys, Reality, Up, series, OP, ív, UP, □, narc
layers.2.mlp	Hindi	53504	agen, Means, uc, Passenger, coil, hist, branch, оч, Means, ifacts
layers.2.mlp	Hindi	70081	retal, udic, fascination, lại, finalize, BSD, eron, repay, FINAL, Targets
layers.3.mlp	Hindi	12962	conta, Aspect, phis, meer, □, acci, strap, probe, liers, 压
layers.3.mlp	Hindi	82959	stown, Tip, SCO, tip, ans, topic, modo, lasts, brush, Stir
layers.3.mlp	Hindi	85531	Depot, aukee, ynn, modo, ITIVE, Rep, rep, ीय, eyh, kuk
layers.3.mlp	Hindi	106101	ради, lda, urse, 赤, venta, під, pta, .ask, urt, 찰
layers.3.mlp	Hindi	128429	chema, alah, Conc, PTY, eil, TRACK, track, aji, segregated, azel
layers.4.mlp	Hindi	37982	erst, lot, Videos, Guru , reme, Bengal , incomplete, Chand , BJP , Kapoor
layers.4.mlp	Hindi	76042	Maharashtra , Gupta , India , Nagar , Uttar , Karnataka , Mumbai , Rao , Goa , Kapoor
layers.4.mlp	Hindi	98694	待, ilded, -Muslim, rick, Expert, senior, presso, /cl, upal, ступ
layers.5.mlp	Hindi	128732	als, oms, linky, Body, Gems, ael, engel, absolute, /body, Mug
layers.7.mlp	Hindi	129521	plugs, waived, orney, ξ, ulaire, aland, sale, LESS, òn, acias
layers.8.mlp	Hindi	13166	vang, oneself, vál, 係, jection, relevance, thereof, 徑, ophis, via
layers.8.mlp	Hindi	17924	Trit, ense, arte, aware, Sense, askell, arti, olt, -CS, Enumerator
layers.8.mlp	Hindi	48133	central, ahkan, ARG, raw, cts, possibly, these, -standing, endent, Browse
layers.8.mlp	Hindi	88795	Queen, 集, entre, Delegate, aper, Directories, 取, aras, eve, Delegate
layers.9.mlp	Hindi	12012	Planet, Planet, planet, ritz, Redirect, alb, societies, Broadcasting, iffs, urv
layers.9.mlp	Hindi	26002	ayi, enga, chip, ndata, .ms, yet, rug, Dram, Manga, fieldValue
layers.9.mlp	Hindi	29761	indir, erville, иться, y, □, himself, House, ville, Shop, elib
layers.9.mlp	Hindi	69906	sth , Sv , Pad , adh , Gand , lok , Pad , amy , prat , Kath
layers.9.mlp	Hindi	77447	Keeper, ista, kee, extras, keeper, imonial, (axis, аря, assin, umble
layers.9.mlp	Hindi	94557	장을, Tu, etsk, parse, Arg, ariate, aren, 交易, hoa, zw
layers.9.mlp	Hindi	119704	esor, erdem, erek, ibli, undry, мак, Í, uncomment, erton, grams