Layer	Lang	Feature ID	Interpretation	Detection				Fuzzing			
				Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.
model.layers.11.mlp	Turkish	38442	The highlighted tokens are predominantly Turkish suffixes and inflectional endings that modify word meaning, indicate possession, plurality, case, tense, or person, as well as some noun and verb roots. These morphological markers are essential for grammatical structure and meaning in Turkish sentences.	0.63	0.41	1.00	0.26	0.68	0.54	0.95	0.38
model.layers.11.mlp	Turkish	55507	The highlighted tokens are predominantly Turkish morphemes, noun and verb roots, and suffixes that form key parts of words, especially those marking tense, plurality, possession, or case. Many are used in constructing complex noun phrases, verb conjugations, or expressing relationships such as agency, time, and location. The activations focus on linguistically significant segments that contribute to the grammatical structure and meaning of sentences.	0.74	0.65	1.00	0.48	0.78	0.73	0.94	0.60
model.layers.11.mlp	Turkish	56369	The highlighted tokens are Turkish verb and noun roots, suffixes, and inflectional endings, often marking tense, person, plurality, or case, as well as forming compound words and participles. These morphemes are essential for word formation and grammatical structure in Turkish.	0.69	0.55	1.00	0.38	0.69	0.56	0.95	0.40
model.layers.11.mlp	Turkish	127127	The highlighted tokens are predominantly proper nouns, especially names of people, places, and groups, often in the context of multicultural or multilingual event descriptions. These tokens frequently appear in various scripts and languages, and are often associated with performances, orchestras, or notable individuals.	0.61	0.55	0.65	0.48	0.55	0.24	0.78	0.14
model.layers.12.mlp	Turkish	40217	The highlighted tokens are Turkish verb roots and suffixes, often marking verb stems, tense, aspect, or person, and are frequently found at the end of words to indicate actions, states, or processes.	0.86	0.84	1.00	0.72	0.90	0.89	0.98	0.82
model.layers.12.mlp	Turkish	45932	The highlighted tokens are predominantly Turkish morphemes, suffixes, and noun or verb stems, often marking grammatical features such as possession, plurality, tense, or case. Many are parts of compound words, inflections, or derivational endings, and frequently appear at the end of words or as part of agglutinative constructions, reflecting the morphological richness and structure of Turkish.	0.79	0.74	0.97	0.60	0.80	0.76	0.97	0.62
model.layers.12.mlp	Turkish	69606	The highlighted tokens are predominantly Turkish verb roots and suffixes, especially those forming various tenses, voices, and moods, as well as participles and nominalizations. There is a strong focus on verb morphology, including derivational and inflectional endings that indicate person, tense, aspect, modality, and negation. These patterns reflect the agglutinative structure of Turkish, where meaning is built up through the sequential addition of suffixes to verb and noun stems.	0.70	0.57	1.00	0.40	0.75	0.68	0.96	0.52
model.layers.12.mlp	Turkish	102905	The important tokens are predominantly Turkish morphemes, suffixes, and stems that form or modify nouns, verbs, and adjectives, often marking tense, person, plurality, comparison, or possession. These tokens are crucial for grammatical structure and meaning in Turkish, highlighting the agglutinative nature of the language where meaning is built up through the addition of multiple suffixes to word roots.	0.77	0.70	1.00	0.54	0.77	0.73	0.89	0.62
model.layers.12.mlp	Turkish	106294	The highlighted tokens are Turkish morphemes, suffixes, or stems that frequently appear at the end or within words, often marking grammatical features such as tense, possession, plurality, or forming nouns and verbs. These segments are crucial for word formation and meaning in Turkish morphology.	0.93	0.93	1.00	0.86	0.92	0.92	0.98	0.86
model.layers.12.mlp	Turkish	112037	The highlighted tokens are predominantly Turkish verb and noun roots, often with attached suffixes, as well as some common noun and adjective stems. These roots frequently appear at the beginning or within words, and are central to the meaning and grammatical structure of the sentences, reflecting the agglutinative nature of Turkish morphology.	0.84	0.82	0.95	0.72	0.83	0.81	0.92	0.72
model.layers.13.mlp	Turkish	3558	The highlighted tokens are predominantly Turkish noun and verb phrases, often marking key semantic units such as actions, roles, attributes, or relationships. These include compound nouns, nominalizations, and verb forms with suffixes indicating tense, possession, or plurality. The tokens frequently appear at phrase or clause boundaries, and often encapsulate the main informational or functional content of the sentence.	0.91	0.90	1.00	0.82	0.91	0.90	1.00	0.82