Layer	Lang	Feature ID	Interpretation	Detection				Fuzzing			
				Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.
model.layers.14.mlp	Turkish	33836	The highlighted tokens are predominantly Turkish suffixes, case endings, and common word stems, as well as function words and frequently used noun or verb forms. These elements are essential for grammatical structure, word formation, and meaning in Turkish sentences, often marking tense, possession, plurality, or case, and sometimes indicating proper nouns or key content words.	0.91	0.90	0.98	0.84	0.89	0.89	0.92	0.86
model.layers.14.mlp	Turkish	43402	The highlighted tokens are predominantly suffixes, inflections, and morphemes in Turkish and related languages, as well as proper nouns, dates, and grammatical particles. These tokens often mark case, possession, plurality, tense, or are part of compound words and names, reflecting the agglutinative structure of the language and the importance of morphological boundaries and named entities in text processing.	0.94	0.94	0.92	0.96	0.91	0.91	0.89	0.94
model.layers.14.mlp	Turkish	51854	The highlighted tokens are predominantly Turkish suffixes, inflections, and root morphemes, especially those involving \"gö\", \"kü\", \"ü\", \"ö\", and other vowel-rich or agglutinative elements, often marking grammatical features like possession, plurality, tense, or forming nouns and adjectives. These tokens are central to Turkish word formation and meaning.	0.93	0.93	0.96	0.90	0.90	0.90	0.94	0.86
model.layers.14.mlp	Turkish	65657	The highlighted tokens are predominantly Turkish morphemes, suffixes, and function words that play key roles in sentence structure, verb conjugation, possession, and meaning. These include verb endings, possessive and case suffixes, conjunctions, and common auxiliary words, reflecting the agglutinative nature of Turkish and the importance of these elements in constructing grammatical and semantically complete sentences.	0.84	0.81	1.00	0.68	0.82	0.79	0.94	0.68
model.layers.14.mlp	Turkish	87172	The highlighted tokens are predominantly Turkish morphemes, roots, and affixes, often marking noun and verb stems, derivational or inflectional endings, and common word fragments. These tokens frequently appear at the beginning or end of words, reflecting the agglutinative structure of Turkish, where meaning is built up through the addition of suffixes and prefixes to roots.	0.90	0.89	1.00	0.80	0.92	0.92	0.96	0.88
model.layers.14.mlp	Turkish	91273	The highlighted tokens are predominantly Turkish morphemes, roots, and suffixes that form or modify nouns, adjectives, and verbs, often marking possession, plurality, tense, or case. These segments frequently appear at the end or within words, reflecting Turkish's agglutinative structure where meaning is built up through the addition of affixes.	0.91	0.90	1.00	0.82	0.93	0.93	0.98	0.88
model.layers.14.mlp	Turkish	106497	The highlighted tokens predominantly correspond to Turkish noun and adjective roots, suffixes, and compound structures, often marking key semantic units such as institutions, official terms, services, locations, and formal processes. There is a strong emphasis on morphological components (roots and suffixes) that form the core meaning of words, especially in administrative, legal, travel, and service-related contexts. These patterns reflect the agglutinative nature of Turkish, where meaning is built up through the combination of roots and multiple suffixes.	0.88	0.87	0.98	0.78	0.92	0.92	0.96	0.88
model.layers.14.mlp	Turkish	108034	The highlighted tokens are predominantly Turkish morphemes, suffixes, and function words that play a key role in grammatical structure, such as denoting possession, plurality, case, tense, or forming compound words and phrases. These tokens often appear at the end of words or as connectors, reflecting the agglutinative nature of Turkish, where meaning and grammatical relationships are built up through the addition of suffixes and short function words.	0.65	0.46	1.00	0.30	0.65	0.49	0.90	0.34
model.layers.15.mlp	Turkish	7049	The highlighted tokens are predominantly Turkish suffixes, inflections, and function words that modify meaning, indicate possession, plurality, tense, or case, as well as common noun and verb roots. These elements are essential for grammatical structure and semantic relationships in Turkish sentences.	0.89	0.88	1.00	0.78	0.90	0.89	0.98	0.82
model.layers.15.mlp	Turkish	12913	The highlighted tokens are predominantly Turkish morphemes, suffixes, and function words, as well as common noun and verb roots. These tokens often mark grammatical relationships, inflections, or frequently used connectors in Turkish, reflecting the language's agglutinative structure and the importance of suffixes for meaning and syntax.	0.83	0.80	0.97	0.68	0.85	0.83	0.97	0.72