Layer	Lang	Feature ID	Interpretation	Detection				Fuzzing			
				Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.
model.layers.15.mlp	Turkish	29918	The highlighted tokens are predominantly Turkish morphemes, suffixes, and word stems that play key grammatical or semantic roles, such as indicating tense, plurality, possession, or forming compound words. These tokens often appear at the ends of words or as part of agglutinative constructions, reflecting the morphological structure of Turkish, where meaning and grammatical function are built up through the addition of suffixes and inflections.	0.75	0.67	1.00	0.50	0.73	0.65	0.93	0.50
model.layers.15.mlp	Turkish	36457	The highlighted tokens are predominantly Turkish morphemes, roots, and suffixes that form the core of word meanings or grammatical functions, such as verb stems, noun roots, and common derivational or inflectional endings. These elements are crucial for constructing and understanding words, indicating tense, possession, plurality, or forming new words, and are often the most semantically or syntactically informative parts of the text.	0.67	0.52	0.95	0.36	0.69	0.62	0.81	0.50
model.layers.15.mlp	Turkish	57867	The highlighted tokens are predominantly Turkish morphemes, roots, and affixes, often marking noun and verb stems, derivational or inflectional endings, and compound word boundaries. These tokens frequently appear at the start or end of words, indicating their importance in Turkish word formation and morphological structure.	0.87	0.85	1.00	0.74	0.87	0.86	0.95	0.78
model.layers.15.mlp	Turkish	67563	The highlighted tokens are predominantly Turkish morphemes, suffixes, and roots that form or modify nouns, verbs, and adjectives, often marking grammatical features such as possession, plurality, tense, or case. Many activations correspond to common Turkish word endings or inflectional affixes, as well as roots of frequently used words, indicating a focus on morphological structure and word formation in Turkish text.	0.84	0.81	1.00	0.68	0.85	0.82	1.00	0.70
model.layers.15.mlp	Turkish	86432	The highlighted tokens are predominantly suffixes, inflections, or short morphemes in Turkish and related languages, often marking grammatical features such as possession, plurality, tense, case, or forming nouns and adjectives. These tokens frequently appear at the end of words and are essential for conveying syntactic and semantic relationships in agglutinative languages.	0.89	0.88	1.00	0.78	0.80	0.81	0.78	0.84
model.layers.15.mlp	Turkish	123739	The highlighted tokens are predominantly short morphemes, syllables, or single letters, often at the beginning of words or as standalone initials, and include both uppercase and lowercase forms. These tokens frequently represent prefixes, roots, or grammatical markers in various languages, and are often found in proper nouns, technical terms, or as part of compound words. The pattern reflects a focus on linguistically meaningful subword units and their role in word formation and structure across multilingual text.	0.55	0.69	0.53	0.98	0.51	0.67	0.51	0.98