Layer	Lang	Feature ID	Interpretation	Detection				Fuzzing			
				Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.
model.layers.13.mlp	Turkish	25033	The highlighted tokens are predominantly Turkish morphemes, suffixes, and noun or verb roots, often marking plurality, possession, tense, or forming compound words. Many are parts of common noun or verb constructions, especially those denoting actions, states, or groupings, and frequently appear at word endings or as inflectional/derivational elements.	0.68	0.53	1.00	0.36	0.69	0.60	0.85	0.46
model.layers.13.mlp	Turkish	34094	The highlighted tokens frequently mark Turkish grammatical suffixes, numerals, time expressions (years, months, days), quantifiers, and common function words. These patterns are typical in Turkish text for expressing dates, quantities, durations, and grammatical relationships.	0.90	0.90	0.94	0.86	0.93	0.93	0.94	0.92
model.layers.13.mlp	Turkish	42039	The highlighted tokens are predominantly Turkish morphemes, suffixes, and common word stems, including noun and verb endings, possessive and case markers, and frequently used roots. These elements are essential for grammatical structure, word formation, and meaning in Turkish, often marking tense, plurality, possession, or case, and are crucial for understanding and generating morphologically rich Turkish text.	0.94	0.94	1.00	0.88	0.96	0.96	0.98	0.94
model.layers.13.mlp	Turkish	63582	The highlighted tokens are predominantly Turkish suffixes and endings that indicate grammatical relationships such as possession, plurality, tense, person, and case, as well as common noun and verb forms. These morphemes are essential for sentence structure and meaning in Turkish, often attached to root words to convey nuanced information.	0.77	0.70	1.00	0.54	0.77	0.70	1.00	0.54
model.layers.13.mlp	Turkish	89356	The highlighted tokens are predominantly Turkish morphemes, roots, and suffixes that form or modify nouns, verbs, and adjectives, often marking grammatical features such as possession, plurality, tense, or case. These segments frequently appear at word boundaries or as inflectional/derivational endings, reflecting the agglutinative structure of Turkish.	0.82	0.78	1.00	0.64	0.85	0.82	1.00	0.70
model.layers.13.mlp	Turkish	94117	The highlighted tokens are primarily Turkish morphemes, compound nouns, and suffixes that form key semantic units such as professions, institutions, services, actions, and descriptors. These tokens often appear at the end of words or as part of compound structures, marking grammatical roles (e.g., possession, plurality, case), or denoting specific domains (e.g., programs, organizations, locations, official documents). The pattern reflects the agglutinative nature of Turkish, where meaning is built up through the addition of suffixes and compound elements.	0.84	0.81	0.97	0.70	0.93	0.93	1.00	0.86
model.layers.13.mlp	Turkish	102015	The highlighted tokens are predominantly Turkish suffixes and inflections that indicate tense, person, plurality, possession, negation, and modality, as well as some common noun and verb roots. These morphological markers are essential for conveying grammatical relationships and meaning in Turkish sentences.	0.64	0.44	1.00	0.28	0.72	0.62	0.96	0.46
model.layers.13.mlp	Turkish	119094	The highlighted tokens are predominantly Turkish suffixes and inflectional endings that attach to word stems to indicate grammatical relationships such as possession, plurality, tense, case, and person. These suffixes are essential for conveying meaning and structure in Turkish sentences, and their activation reflects their importance in determining the function and relationship of words within the text.	0.66	0.49	1.00	0.32	0.64	0.46	0.94	0.30
model.layers.13.mlp	Turkish	119659	The highlighted tokens are predominantly Turkish morphemes, suffixes, and function words that play a key role in sentence structure, verb conjugation, possession, and case marking, as well as common connectors and pronouns. These elements are essential for expressing grammatical relationships and meaning in Turkish sentences.	0.83	0.80	1.00	0.66	0.83	0.80	1.00	0.66
model.layers.14.mlp	Turkish	22302	The highlighted tokens are predominantly Turkish morphemes, suffixes, and stems that form the core of word meanings, verb conjugations, and noun/adjective derivations. These include tense, person, plurality, possession, and case markers, as well as common roots and affixes that are essential for grammatical structure and semantic content in Turkish sentences. The activations focus on the morphological building blocks that determine the function and meaning of words within the sentence.	0.74	0.65	1.00	0.48	0.76	0.68	1.00	0.52
model.layers.14.mlp	Turkish	30275	The highlighted tokens are predominantly Turkish suffixes, inflections, and function words that modify meaning, indicate grammatical relationships, or form noun and verb phrases. These include case endings, possessives, plural markers, verb conjugations, and common connectors, reflecting the agglutinative structure of Turkish and the importance of morphological units in sentence construction.	0.76	0.68	1.00	0.52	0.78	0.72	1.00	0.56