

Layer	Lang	Feature ID	Interpretation	Detection				Fuzzing			
				Accuracy	F1 score	Precision	Recall	Accuracy	F1 score	Precision	Recall
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	0.685	0.527	0.98	0.51	0.667	0.505	0.98