Layer	Lang	Feature ID	Interpretation	Detection				Fuzzing			
				Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.
model.layers.14.mlp	Thai	11744	The highlighted tokens are primarily function words, pronouns, particles, and common morphemes in Thai, often marking grammatical relationships, sentence structure, or serving as connectors. There is a strong emphasis on words that indicate possession, agency, time, location, and conjunctions, as well as frequent use of polite particles and verb auxiliaries. These tokens are essential for the syntactic and semantic coherence of Thai sentences.	0.73	0.66	0.90	0.52	0.76	0.68	1.00	0.52
model.layers.14.mlp	Thai	22779	The highlighted segments are primarily Thai words, phrases, or morphemes that serve as meaningful units within sentences, including nouns, verbs, particles, and function words. These segments often correspond to syntactic or semantic boundaries, such as names, actions, objects, or grammatical markers, and sometimes include foreign words or transliterations. The activations tend to focus on morphemes or syllables that carry core meaning or grammatical function within the context.	0.88	0.86	1.00	0.76	0.86	0.84	0.97	0.74
model.layers.14.mlp	Thai	33861	The highlighted tokens are primarily Thai morphemes, syllables, or short words that serve as grammatical markers, connectors, or key semantic units within sentences. These include function words, affixes, and core content words that are essential for sentence structure and meaning, such as those indicating possession, agency, location, comparison, or action. The pattern reflects the importance of these units in Thai syntax and information flow, often marking relationships between clauses, specifying entities, or denoting actions and attributes.	0.99	0.99	1.00	0.98	0.97	0.97	1.00	0.94
model.layers.14.mlp	Thai	34721	The highlighted tokens are overwhelmingly function words, pronouns, and common morphemes or syllables that serve as grammatical connectors in Thai, such as markers for possession, location, time, or subject/object reference. There is a strong emphasis on high-frequency, short tokens that are essential for sentence structure and meaning, including prefixes, suffixes, and particles that modify or clarify the main content words. These tokens are crucial for the cohesion and flow of Thai sentences.	0.66	0.49	1.00	0.32	0.57	0.47	0.61	0.38
model.layers.14.mlp	Thai	54904	The highlighted tokens are primarily morphemes, affixes, or short words that serve as grammatical connectors, markers of tense, aspect, or case, and components of compound words in Thai. They often appear at the boundaries of words or phrases, contributing to the structure and meaning of sentences by indicating relationships, possession, comparison, or forming part of proper nouns and technical terms.	0.87	0.85	1.00	0.74	0.87	0.87	0.89	0.84
model.layers.14.mlp	Thai	62719	The highlighted tokens are primarily Thai morphemes, syllables, or short words that function as grammatical markers, noun or verb roots, or affixes. They often appear at the boundaries of phrases or as key components in compound words, and are frequently associated with high-importance content such as actions, conditions, or objects within sentences. The pattern reflects a focus on semantically or syntactically significant units in Thai text.	0.70	0.57	1.00	0.40	0.69	0.55	1.00	0.38
model.layers.14.mlp	Thai	73972	The pattern highlights the importance of single or paired consonant-vowel tokens, especially those involving \"ต\" (Thai), \"ท\" (Thai), and \"त\" (Hindi), which frequently appear at morpheme or word boundaries, often as part of grammatical constructions, inflections, or compound words in Thai and Hindi text. These tokens are crucial for forming or modifying meaning within words and phrases.	0.78	0.76	0.85	0.68	0.81	0.80	0.83	0.78
model.layers.14.mlp	Thai	101719	The highlighted tokens are primarily function words, particles, and morphemes that serve grammatical roles in Thai, such as marking tense, aspect, conjunctions, pronouns, and case. There is a strong emphasis on connectors, sentence structure markers, and suffixes that modify meaning or indicate relationships between clauses and entities. These tokens are essential for the syntactic and semantic cohesion of Thai sentences.	0.85	0.82	1.00	0.70	0.82	0.79	0.94	0.68
model.layers.15.mlp	Thai	42804	The highlighted tokens are primarily morphemes, syllables, or short word fragments in Thai and English, often marking key semantic units such as nouns, verbs, locations, or grammatical particles. These segments frequently appear at the boundaries of words or phrases, and are important for identifying meaning, structure, or named entities within multilingual or code-mixed text.	0.70	0.64	0.79	0.54	0.54	0.60	0.53	0.70