

Feature 118169

Activation Range

Language: Thai	0.085-0.179	0.179-0.273	0.273-0.366	0.366-0.46	0.46-0.554	0.554-0.648	0.648-0.742
Model: meta-llama/Llama-3.2-1B	0.742-0.835	0.835-0.929	0.929-1.023				
Layer: model.layers.5.mlp							
SAE Model: EleutherAI/sae-Llama-3.2-1B-131k							
Selected Token Probability: 0.819							
Entropy: 0.397							

Interpretation

"The highlighted tokens correspond to the word \"Thailand\" and its demonyms or derivatives, as well as related country names, written in various languages and scripts. The pattern is the consistent identification of the country \"Thailand\" or its linguistic equivalents across multilingual contexts."

Score Type	Accuracy	Precision	Recall	F1 score	TPR	TNR	FPR	FNR
detection	0.54	1.0	0.08	0.148	0.08	1.0	0.0	0.92
fuzz	0.54	1.0	0.08	0.148	0.08	1.0	0.0	0.92

Thai

#examples: [('xnli', 1000), ('flores', 997)]

xnli-183. < begin_of_text >เรา ได้ ไป ใน ทริป ที่ เรา ได้ อามน้ำ ใน ลำธาร นอกจาก การอามน้ำ ใน ลำธาร แล้ว เรา ก็ ไป สป่า และ ชาว นา ด้วย
xnli-277. < begin_of_text >ยัง มี พิพิธภัณฑ์ โบราณคดี ที่ แสดง ศิลปวัตถุ ที่ เก่า กว่า รวมถึง ตัวอย่าง ของ เครื่องปั้นดินเผา เครื่องปั้นดินเผา พิพิธภัณฑ์ ว้าง มาก และ ไม่มี อะไร อยู่ ใน นั้น เลย
xnli-127. < begin_of_text >ใน หมู่ พวก นี้ คือ วัง อีรุ แดง ซึ่ง ตอนนี บ้าน พิพิธภัณฑ์ patan (เนปาล ' s ที่ ดี ที่สุด และ พิพิธภัณฑ์ หั่นสมัย ที่สุด) และ หั่นหน้า เข ่า วัง ตรงข้าม อีรุ แดง พลาซ่า แปด วัต ของ สไตส์ และ ขนาด ต่างๆ พิพิธภัณฑ์ patan อยู่ริมถนน จาก วัง อีรุ แดง ราชบุรี
xnli-824. < begin_of_text >แคล ผลัก จะ ' daan นอกจาก มือ ที่ มีพลัง เดียว Ca ' daan ถูก ผลัก ออก ไป และ เคาะ ฟัน โดย มือ ที่ มีพลัง ของ คา ล

Text Examples for Each Interval

interval 1

Range: 0.929-1.023

#examples: 1

flores-441. <|begin_of_text|>Nel XVIII secolo la Cambogia fu invasa più volte dai thailandesi, che nel 1772 distrussero la capitale Phnom Phen.

interval 2

Range: 0.835-0.929

#examples: 5

flores-663. < begin_of_text >São Francisco é também um dos melhores lugares do país no que diz respeito a outras opções de culinária asiática: comida coreana, tailandesa, indiana e japonesa.
flores-663. < begin_of_text >San Francisco ist auch einer der landesweit besten Orte für andere asiatische Küche: Koreanisch, Thailandisch, Indisch und Japanisch.
flores-440. < begin_of_text >През 18 -ти век се оказва, че Камбоджа е притисната между двама могъщи съседи, Тайланд и Виетнам.
flores-440. < begin_of_text >EL Reino de Camboya quedó acorralado entre dos poderosos vecinos durante el siglo XVIII. Ellos eran Tailandia y Vietnam.
flores-440. < begin_of_text >Nel corso del XVIII secolo la Cambogia si trovava stretta tra due vicini molto potenti, la Thailandia e il Vietnam.

interval 3

Range: 0.742-0.835

#examples: 7

flores-440. <|begin_of_text|>18. asırda Kamboçya kendisini iki güçlü komşu olan Tayland ve Vietnam arasında buldu.