

Feature 61383

Language: Vietnamese  
Model: meta-llama/Llama-3.2-1B  
Layer: model.layers.9.mlp  
SAE Model: EleutherAI/sae-Llama-3.2-1B-131k  
Selected Token Probability: 0.881  
Entropy: 0.211

Activation Range

0.097-0.368	0.368-0.639	0.639-0.911	0.911-1.182	1.182-1.453	1.453-1.724	1.724-1.995
1.995-2.267	2.267-2.538	2.538-2.809				

Interpretation

"The highlighted tokens correspond to names of Cambodian places and landmarks, especially those related to Tonle Sap, Phnom Krom, Angkor, and Siem Reap, often appearing in multiple languages and scripts, with activations on both full names and their constituent parts."

Score Type	Accuracy	Precision	Recall	F1 score	TPR	TNR	FPR	FNR
detection	0.51	1.0	0.02	0.039	0.02	1.0	0.0	0.98
fuzz	0.51	1.0	0.02	0.039	0.02	1.0	0.0	0.98

Vietnamese

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

- xnli-298. <|begin\_of\_text|>Bình luận về các vấn đề về thực thi và câu hỏi được nuôi dưỡng bởi công nghiệp bị kiểm soát được gửi bởi epa trong bốn câu hỏi quy tắc và trả lời các tài liệu . Tất cả các vấn đề và câu hỏi được nuôi dưỡng bởi các công nghiệp bị kiểm soát được đưa ra bởi epa .
- xnli-931. <|begin\_of\_text|>Hôm nay nó nhà một bảo tàng nhỏ ? . Nó chứa một viện bảo tàng nhỏ .
- xnli-568. <|begin\_of\_text|>Họ thậm chí còn có thể ngồi cạnh tài xế . Chỗ ngồi bên cạnh tài xế cung cấp những quan điểm tốt nhất .
- xnli-906. <|begin\_of\_text|>Nghe có vẻ như đó là một ý tưởng rất hay đấy . Nghe có vẻ là một ý tưởng khủng khiếp .

Text Examples for Each Interval

interval 1

Range: 2.538-2.809  
#examples: 9

- flores-690. <|begin\_of\_text|>A atmosfera sombria do templo e a vista do lago Tonle Sap fazem a subida à colina valer a pena.
- flores-441. <|begin\_of\_text|>Durante el siglo XVIII Camboya sufrió varias invasiones de los tailandeses, quienes, en 1772, destruyeron Phnom Phen.
- flores-690. <|begin\_of\_text|>Tapınağın kasvetli atmosferi ve Tonle Sap gölü manzarası, tepeye tırmanmayı zahmete değer bir hale getiriyor.
- flores-690. <|begin\_of\_text|>寺阴的围和洞里萨湖（Tonle Sap）的景色，让人感到到山顶是值得的。
- flores-690. <|begin\_of\_text|>Vale la pena scalare la collina per godersi l'atmosfera tenebrosa di questo tempio e il panorama sul lago Tonle Sap.
- flores-692. <|begin\_of\_text|>Bạn cần vé vào Angkor để vào đền, vậy nên đừng quên mang theo hộ chiếu của bạn khi đi đến Tonle Sap.
- flores-692. <|begin\_of\_text|>需要购买吴哥门票才能进入圣殿，所以前往洞里萨湖（Tonle Sap）时别忘了带上护照。
- flores-690. <|begin\_of\_text|>The gloomy atmosphere of the temple and the view over the Tonle Sap lake make the climb to the hill worthwhile.
- flores-692. <|begin\_of\_text|>The Angkor Pass is needed to enter the temple so do not forget to bring your passport along when heading to Tonle Sap.

interval 2

Range: 2.267-2.538  
#examples: 18

- flores-690. <|begin\_of\_text|>La atmósfera sombría del templo y la vista del lago Tonle Sap hacen que el ascenso por la colina tenga su recompensa.
- flores-441. <|begin\_of\_text|>18. yüzyılda birçok kez Kamboçya'yı işgal eden Taylandlılar 1772'de Phnom Phen'i yerle bir etmiştir.