Feature 67845

Language: Korean
Model: meta-llama/Llama-3.2-1B
Layer: model.layers.12.mlp
SAE Model: EleutherAI/sae-Llama-3.2-1B-131k
Selected Token Probability: 0.136
Entropy: 0.861

Activation Range

0.071-0.804 0.804-1.537 1.537-2.271 2.271-3.004 3.004-3.737 3.737-4.47 4.47-5.203 5.203-5.937 5.937-6.67 6.67-7.403

Interpretation

"The text consistently centers on King Sejong and the Joseon Dynasty, the invention of the Hangeul (Hangul) alphabet, and the name \"Hunmin Jeongeum,\" with key tokens marking royal titles, dynastic references, invention dates (especially 1444 and 1418–1450), and the names of the alphabet and its inventor, across multiple languages."

Score Type	Accuracy	Precision	Recall	F1 score	TPR	TNR	FPR	FNR
detection	0.53	1.0	0.06	0.113	0.06	1.0	0.0	0.94
fuzz	0.53	1.0	0.06	0.113	0.06	1.0	0.0	0.94

Korean

#examples: [('paws-x', 994), ('flores', 995)]

paws-x-566. <|begin_of_text|>경기 1 : Sano Naoki가 Kakihara Masahito를 물리 ��다.

paws-x-462. <|begin_of_text|>그녀는 그에게 크게 매료되어 그를 성적 관계로 유혹하려고했지만 Hanuvant Singh은 종교적인 생각으로 근친 상간에 가지 않았습니다.

paws-x-555. <|begin_of_text|>오늘 밤, Muzong 황제는 죽고, Li Zhan는 (황제 Jingzong로) 왕위에 앉았다.

paws-x-155. <|begin_of_text|>Neyab (또한 Neyab로 로마자 표기)은 Esfarayen 카운티, 북부 Khorasan 주,이란, Bam 및 Safiabad District, Safiabad Rural District에있는 마을입니다.

Text Examples for Each Interval

interval 1

Range: 6.67-7.403
#examples: 1

flores-386. <|begin_of_text|>Il re Sejong fu il quarto re della dinastia Joseon e uno dei sovrani più stimati.

interval 2

Range: 5.937-6.67
#examples: 7

flores-385. <|begin_of_text|>한글은 유일하게 일상적으로 ��리 사용하기 위해 특별히 고안된 글자이다. 한글은 세종대왕(1418~1450) 때인 1444년에 발명되었다.

flores-385. <|begin_of_text|>Lo hangeul è l'unico alfabeto inventato intenzionalmente per l'uso quotidiano da parte del popolo. Fu ideato durante il regno del Re Sejong (1418-1450) nel 1444.

flores-386. <|begin_of_text|>Sejong le Grand fut le quatrième roi de la dynastie Joseon et demeure l'un des souverains coréens les plus respectés.

flores-385. <|begin_of_text|>ハングルは、日常的に使われている唯一の字母です。字母は世宗時代(1418~1450)の1444年に発明されました。

flores-385. <|begin_of_text|>El alfabeto coreano es el único diseñado en forma deliberada que aún se utiliza a diario popularmente. Se inventó en 1444, durante el reinado de Sejong (1418 a 1450).

flores-386. <|begin_of_text|>King Sejong was the fourth king of the Joseon Dynasty and is one of the most highly regarded.

flores-385. <|begin_of_text|>ฮันกึลเป็นอักษรที่ประดิษฐ์ขึ้นโดยเจตนาเพียงชุดเดียวที่นิยมใช้ในชีวิตประจำวัน ชุดอักษรนี้ประดิษฐ์ขึ้นในปีค.ศ. 1444 ในรัชสมัยของกษัตริย์เซจง (ค.ศ. 1418 - 1450)