Feature 5889

Language: English
Model: meta-llama/Llama-3.2-1B
Layer: model.layers.15.mlp
SAE Model: EleutherAI/sae-Llama-3.2-1B-131k
Selected Token Probability: 0.183
Entropy: 1.901

Activation Range

| 0.213-1.443 | 1.443-2.673 | 2.673-3.904 | 3.904-5.134 | 5.134-6.364 | 6.364-7.594 | 7.594-8.824 |

| 8.824-10.055 | 10.055-11.285 | 11.285-12.515 |

Interpretation

"The highlighted tokens frequently correspond to structural elements in formal or factual writing, such as numbers, dates, conjunctions, prepositions, punctuation, and proper nouns, which are often used to organize, enumerate, or specify information in lists, references, or descriptive passages."

Score Type	Accuracy	Precision	Recall	F1 score	TPR	TNR	FPR	FNR
detection	0.7	0.727	0.64	0.681	0.64	0.76	0.24	0.36
fuzz	0.64	0.625	0.7	0.66	0.7	0.58	0.42	0.3

English

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

xnli-720. <|begin_of_text|>Comments on classified or restricted reports should be transmitted in the manner agreed to by GAO and the agency . There are specific transmission methodologies for classified reports .

xnli-575. <|begin_of_text|>Analysis of this information revealed that no two of these organizations defined the design review process and its various elements in exactly the same manner . Many organizations have a radically different approach to their design review process .

xnli-36. <|begin_of_text|>there 's a uh a couple called um oh i 'm going to forgot his name now uh Dirkson I can 't remember their name

xnli-589. <|begin_of_text|>i 've had a chance to i had a chance to terminate and uh L O A and back to L O A I never had any opportunity to end LOA .

xnli-732. <|begin_of_text|>Advertisement , replied Tuppence promptly . Tuppence replied right away .

xnli-345. <|begin_of_text|>2) House Judiciary Committee Chairman Henry Hyde said he has begun to examine impeachment procedures in the event they are justified . Henry Hyde is a new house judiciary .

xnli-650. <|begin_of_text|>(5) (A) The term baseline heat input means , except under subpart 1 of part B and section 407 , the average annual heat input used by a unit during the three years in which the unit had the highest heat input for the period 1997 through 2001 . The baseline heat input is the lowest annual heat input over 3 years .

xnli-177. <|begin_of_text|>okay pro football i like two teams one the New York Giants and the second is the Raiders The New York Giants and Raiders are my favorite teams in football .

xnli-550. <|begin_of_text|>so i i know the people at TI who are doing this and i heard about it so i called them and ask if i could could participate and uh I don 't know anyone at TI and that 's okay because I don 't want to participate anyway .

xnli-618. <|begin_of_text|>1.7 , formerly methodology transfer paper 7 . It used to be paper 7 .

xnli-943. <|begin_of_text|>Also the Treasury requires departments to disclose instances of irregular expenditures arising from erroneous benefit awards and fraud by claimants . Departments are typically reluctant to disclose mistakes in the awarding of benefits .

xnli-478. <|begin_of_text|>Breyiana Breyiana look at me please just a second be quite please i can not hear i will talk to you in just a second go ahead so she went to school out here Please be quiet for a second ; I cannot hear and will talk to you in a second .

xnli-119. <|begin_of_text|>In any case--it 's over ." It is over .

xnli-301. <|begin_of_text|>Inside , a huge mountain of a man , even bigger than Thorn , fought with a huge two-handed axe . The man was extremely large .

xnli-54. <|begin_of_text|>Analyzing Postal Service accounts for depreciation , fuel , and maintenance for city delivery carriers , we have estimated the average city delivery vehicle cost per route . Driving cost estimates can be averaged will sufficient data .

xnli-13. <|begin_of_text|>Thebes held onto power until the 12th Dynasty , when its first king , Amenemhet Iwho reigned between 1980 1951 b.c. established a capital near Memphis . The capital near Memphis lasted only half a century before its inhabitants abandoned it for the next capital