**Phi-2 2.7B**
**phi 3 3.8b**

**No Fine-Tuning - Validation Testing without training**

Perplexity on WikiText2 for phi–2
Base - 9.7092
0.2 sparsity - 10.0805
0.4 sparsity - 11.5489
0.6 sparsity - 26.6037
0.8 sparsity - 130250.2734
0.5 sparsity - 2:4 structured - 29.4844

Perplexity on WikiText2 for phi–3
Base - 92.2728
0.2 sparsity - 99.3140
0.4 sparsity - 126.4848
0.6 sparsity - 154.3939
0.8 sparsity - 367495.1875
0.5 sparsity - 2:4 structured - 160.1343

Use visualization instead of tables and no perplexity for dataset validations
Use bar chart or histograms
Write down the vulnerabilities and analysis of each datasets

CommonsenseQA

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-2 | 57.49 | 57.60 | 57.49 | 57.48 |
| 0.2 sp | 56.02 | 56.08 | 56.02 | 55.99 |
| 0.4 sp | 49.63 | 49.73 | 49.63 | 49.63 |
| 0.6 sp | 36.04 | 36.15 | 36.04 | 36.02 |
| 0.8 sp | 17.85 | 17.86 | 17.85 | 17.84 |
| 2:4 - 0.5 sp | 36.20 | 36.31 | 36.20 | 36.22 |

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-3 mini | 48.40 | 48.54 | 48.40 | 48.45 |
| 0.2 sp | 47.58 | 47.68 | 47.58 | 47.61 |
| 0.4 sp | 43.49 | 43.52 | 43.49 | 43.50 |
| 0.6 sp | 32.27 | 32.15 | 32.27 | 32.18 |
| 0.8 sp | 15.23 | 15.23 | 15.23 | 15.23 |
| 2:4 - 0.5 sp | 31.45 | 31.43 | 31.45 | 31.43 |

CosmosQA

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-2 | 30.95 | 30.97 | 30.95 | 30.96 |
| 0.2 | 34.51 | 34.50 | 34.51 | 34.49 |
| 0.4 | 34.47 | 34.45 | 34.47 | 34.45 |
| 0.6 | 30.32 | 30.30 | 30.32 | 30.30 |
| 0.8 | 19.13 | 19.12 | 19.13 | 19.12 |
| 2:4 - 0.5 sp | 30.52 | 30.53 | 30.52 | 30.53 |

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-3 | 31.93 | 31.94 | 31.93 | 31.93 |
| 0.2 | 32.33 | 32.34 | 32.33 | 32.33 |
| 0.4 | 37.22 | 37.21 | 37.22 | 37.21 |
| 0.6 | 28.31 | 28.28 | 28.31 | 28.29 |
| 0.8 | 20.03 | 20.03 | 20.03 | 20.03 |
| 2:4 - 0.5 sp | 28.68 | 28.67 | 28.68 | 28.67 |

LogiQA

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-2 | 21.20 | 28.35 | 21.20 | 20.63 |
| 0.2 | 20.74 | 27.99 | 20.74 | 20.27 |
| 0.4 | 19.66 | 24.77 | 19.66 | 18.82 |
| 0.6 | 18.74 | 23.98 | 18.74 | 18.12 |
| 0.8 | 17.97 | 21.59 | 17.97 | 18.24 |
| 2:4 - 0.5 sp | 19.05 | 23.44 | 19.05 | 18.12 |

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-3 | 21.81 | 30.76 | 21.81 | 21.02 |
| 0.2 | 21.81 | 31.15 | 21.81 | 21.27 |
| 0.4 | 21.04 | 29.43 | 21.04 | 20.77 |
| 0.6 | 20.28 | 27.54 | 20.28 | 19.14 |
| 0.8 | 18.74 | 22.31 | 18.74 | 18.91 |
| 2:4 - 0.5 sp | 20.89 | 26.80 | 20.89 | 20.64 |

ReClor

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-2 | 22.00 | 22.27 | 22.00 | 22.07 |
| 0.2 | 20.40 | 20.72 | 20.40 | 20.49 |
| 0.4 | 20.80 | 21.15 | 20.80 | 20.92 |
| 0.6 | 20.60 | 20.96 | 20.60 | 20.71 |
| 0.8 | 17.00 | 17.20 | 17.00 | 17.08 |
| 2:4 - 0.5 sp | 19.00 | 19.15 | 19.00 | 18.98 |

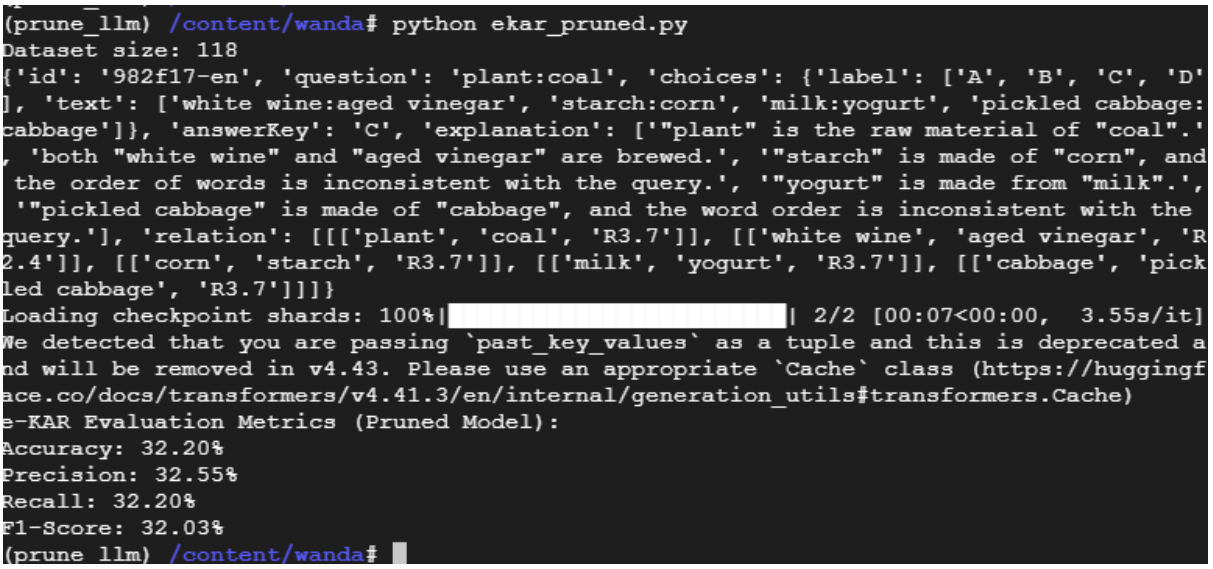
| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-3 | 22.60 | 22.82 | 22.60 | 22.69 |
| 0.2 | 24.00 | 24.10 | 24.00 | 24.04 |
| 0.4 | 22.80 | 23.32 | 22.80 | 23.01 |
| 0.6 | 20.80 | 21.18 | 20.80 | 20.91 |
| 0.8 | 17.00 | 17.25 | 17.00 | 17.10 |
| 2:4 - 0.5 sp | 21.40 | 21.78 | 21.40 | 21.56 |

E-KAR

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-2 | 33.05 | 32.84 | 33.05 | 32.08 |
| 0.2 | 33.90 | 34.16 | 33.90 | 33.47 |
| 0.4 | 38.14 | 38.69 | 38.14 | 38.19 |
| 0.6 | 33.90 | 33.99 | 33.90 | 33.63 |
| 0.8 | 27.97 | 28.06 | 27.97 | 27.99 |
| 2:4 - 0.5 sp | 32.20 | 32.55 | 32.20 | 32.03 |

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Phi-3 | 37.29 | 39.27 | 37.29 | 37.48 |
| 0.2 | 38.98 | 40.83 | 38.98 | 39.00 |
| 0.4 | 40.68 | 43.75 | 40.68 | 40.97 |
| 0.6 | 26.27 | 26.59 | 26.27 | 26.22 |
| 0.8 | 26.27 | 26.31 | 26.27 | 26.22 |
| 2:4 - 0.5 sp | 33.90 | 33.93 | 33.90 | 33.77 |

Example :


(prune_llm) /content/wanda# python ekar_pruned.py
Dataset size: 118
{'id': '982f17-en', 'question': 'plant:coal', 'choices': {'label': ['A', 'B', 'C', 'D'
], 'text': ['white wine:aged vinegar', 'starch:corn', 'milk:yogurt', 'pickled cabbage:
cabbage']}, 'answerKey': 'C', 'explanation': ['"plant" is the raw material of "coal".'
, 'both "white wine" and "aged vinegar" are brewed.', '"starch" is made of "corn", and
the order of words is inconsistent with the query.', '"yogurt" is made from "milk".',
'"pickled cabbage" is made of "cabbage", and the word order is inconsistent with the
query.'], 'relation': [[['plant', 'coal', 'R3.7']], [['white wine', 'aged vinegar', 'R
2.4']], [['corn', 'starch', 'R3.7']], [['milk', 'yogurt', 'R3.7']], [['cabbage', 'pick
led cabbage', 'R3.7']]]}
Loading checkpoint shards: 100%| | 2/2 [00:07<00:00, 3.55s/it]
We detected that you are passing `past_key_values` as a tuple and this is deprecated a
nd will be removed in v4.43. Please use an appropriate `Cache` class (https://huggingf
ace.co/docs/transformers/v4.41.3/en/internal/generation_utils#transformers.Cache)
e-KAR Evaluation Metrics (Pruned Model):
Accuracy: 32.20%
Precision: 32.55%
Recall: 32.20%
F1-Score: 32.03%
(prune_llm) /content/wanda#