

DATASET - cais/mmlu - 120 cases for inference

Llama 8B 3.1 instruct model

Used Together.ai for inference

- 1.)0 SHOT => evaluated on all 120 to get the accuracy 0.397
- 2.)COT => Accuracy of 0.419
- 3.)ReACT => Accuracy of 0.425

Phi 3.5 mini instruct model

Together.ai model for phi was not available . So used Hugging face model . Due to restraint in resources , Evaluated on 20 random problems.

- 1.)0 SHOT => 0.17
- 2.)COT =>0.229
- 3.)ReACT => 0.243

Gemma 2B

Together.ai model available

- 1.)0 SHOT => 0.14
- 2.)COT =>0.197
- 3.)ReACT => 0.214