Comparison of Primary Metric Across Runs

Exact-Match accuracy on GSM8K dev split

Note: All runs have zero accuracy
(Model may not have converged)

0.0000

0.0000

Proposed Method
Baseline/Comparative

proposed-iter1-Qwen3-0.6B-gsm8k

comparative-1-iter1-Qwen3-0.6B-gsm8k

Run ID