
Supplementary Material for Group-MATES

Anonymous Author(s)

Affiliation

Address

email

1 A Rollout Length

Table 1: Results on DCLM 1B-1x setting with different rollout length T .

T	COMMONSENSE REASONING (3 tasks)	LANGUAGE UNDERSTANDING (6 tasks)	READING COMPREHENSION (3 tasks)	SYMBOLIC PROBLEM SOLVING (5 tasks)	WORLD KNOWLEDGE (5 tasks)	CORE (22 tasks)
2	0.36687	0.40461	0.22536	0.20110	0.30817	0.30685
20	0.36549	0.39901	0.21265	0.20501	0.30081	0.30262
10 (OURS)	0.36997	0.39744	0.23922	0.20250	0.30793	0.30747

2 In Table 1, we investigate the effect of varying the rollout length T in Group-MATES, considering
3 values of 2, 10, and 20. We observe that $T = 10$ performs slightly better than $T = 2$, suggesting that
4 a moderate increase in rollout length enables our relational data influence model to better capture
5 long-term effects. However, increasing T to 20 results in a performance decline, likely due to
6 the increased complexity of modeling combinatorial effects over longer trajectories, which incurs
7 additional challenges for the relational data influence model. These results indicate the importance
8 of selecting an appropriate rollout length that sufficiently reflects group-level data influence while
9 remaining tractable for the relational data influence model to learn effectively.