

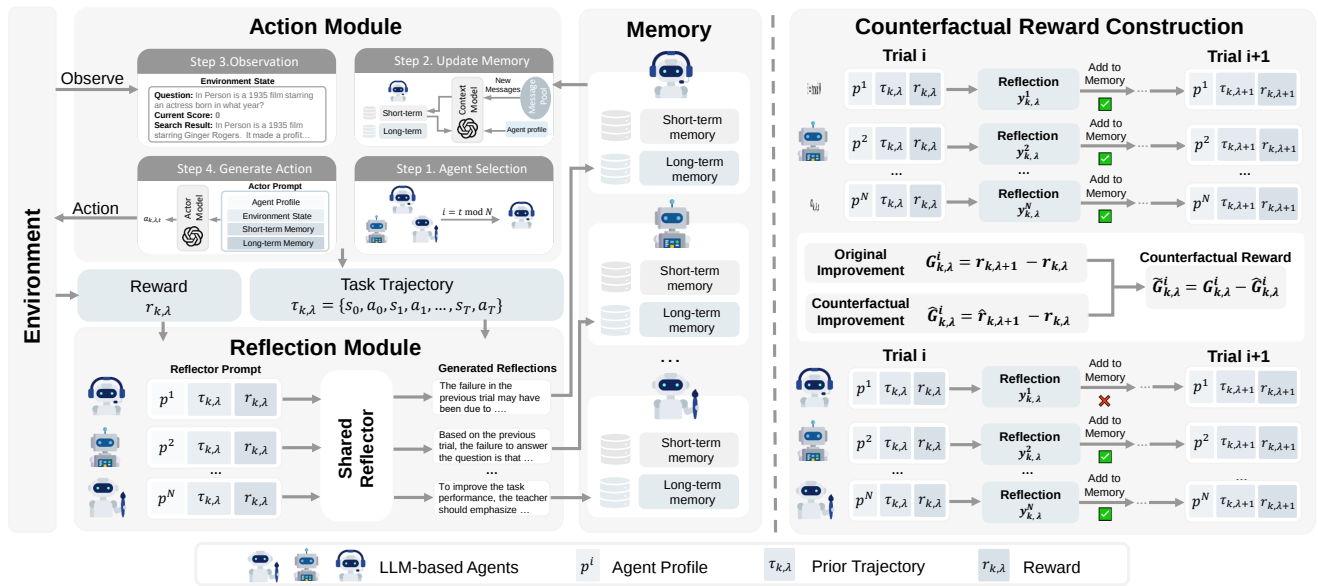


Figure 1: The overview of our proposed COPPER.

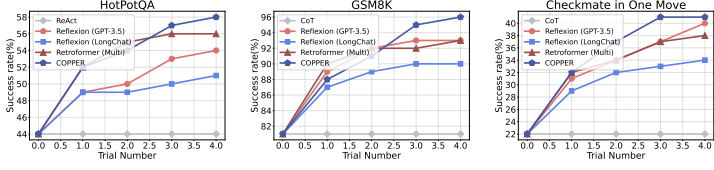


Figure 2: Main results (added Multi-Retroformer baseline).

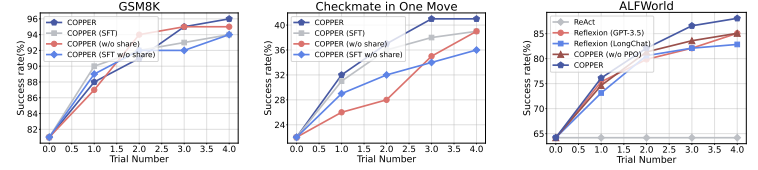


Figure 3: Effectiveness of shared reflector

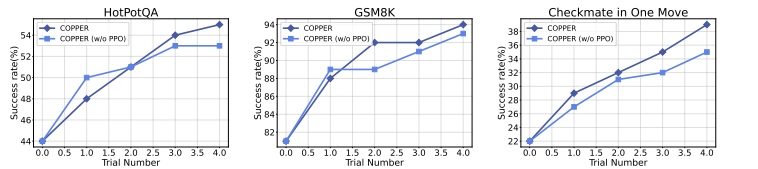
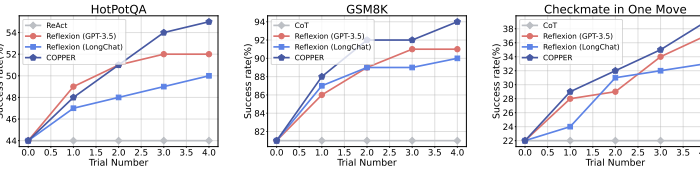


Figure 4: Main results on GSM8K and Checkmate in One Move.



Figure 5: Main results on ALFWorld.

(a) Performance of COPPER against other baselines.

(b) Comparison between SFT and RLHF.

Figure 5: Main results under non-profile setting.

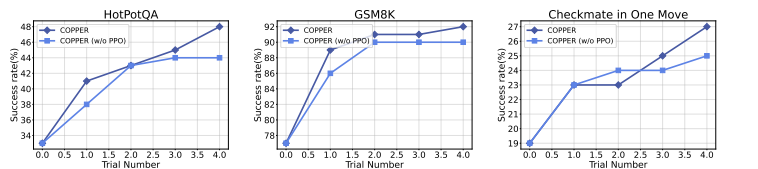
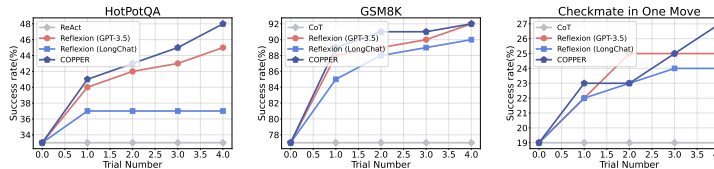


Figure 6: Main results without other agents' trajectories.

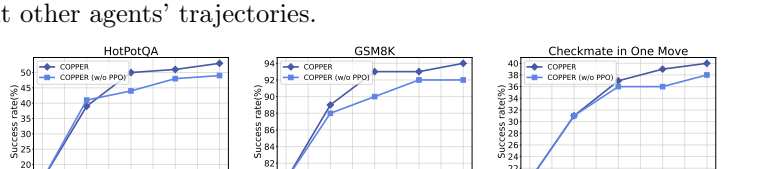
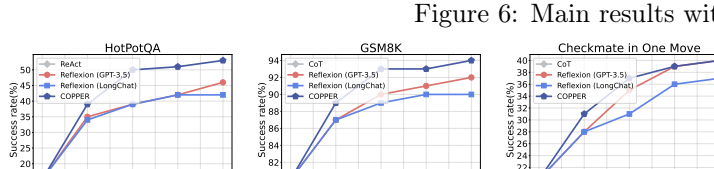


Figure 7: Main results when predicting other agents' information with a proxy model.

Table 1: Statistics of training data generated by original LongChat. We show the total data volume on the left, the number of positive examples on the middle, and the number of negative examples on the right side.

Dataset	Episode Difference Reward				Counterfactual Reward			
	all	agent_0	agent_1	agent_2	all	agent_0	agent_1	agent_2
HotPotQA	8714/1946/1240	4357/973/620	4357/973/620	-/-/-	8714/2137/1330	4357/1065/686	4357/1072/644	-/-/-
GSM8K	3147/852/0	1049/284/0	1049/284/0	1049/284/0	3147/871/417	1049/138/145	1049/128/131	1049/132/141
Checkmate.	17466/906/0	5822/302/0	5822/302/0	5822/302/0	17466/682/461	5822/234/165	5822/223/137	5822/228/159

Table 2: Statistics of training data generated by LongChat fine-tuned with SFT. We show the total data volume on the left, the number of positive examples on the middle, and the number of negative examples on the right side.

Dataset	Episode Difference Reward				Counterfactual Reward			
	all	agent_0	agent_1	agent_2	all	agent_0	agent_1	agent_2
HotPotQA	8822/2218/1410	4411/1109/705	4411/1109/705	-/-/-	8822/1679/1985	4411/855/1027	4411/824/958	-/-/-
GSM8K	3204/789/0	1068/263/0	1068/263/0	1068/263/0	3204/384/409	1068/129/143	1068/128/132	1068/127/134
Checkmate.	17892/591/0	5964/197/0	5964/197/0	5964/197/0	17892/412/444	5964/138/147	5965/138/169	5966/136/128

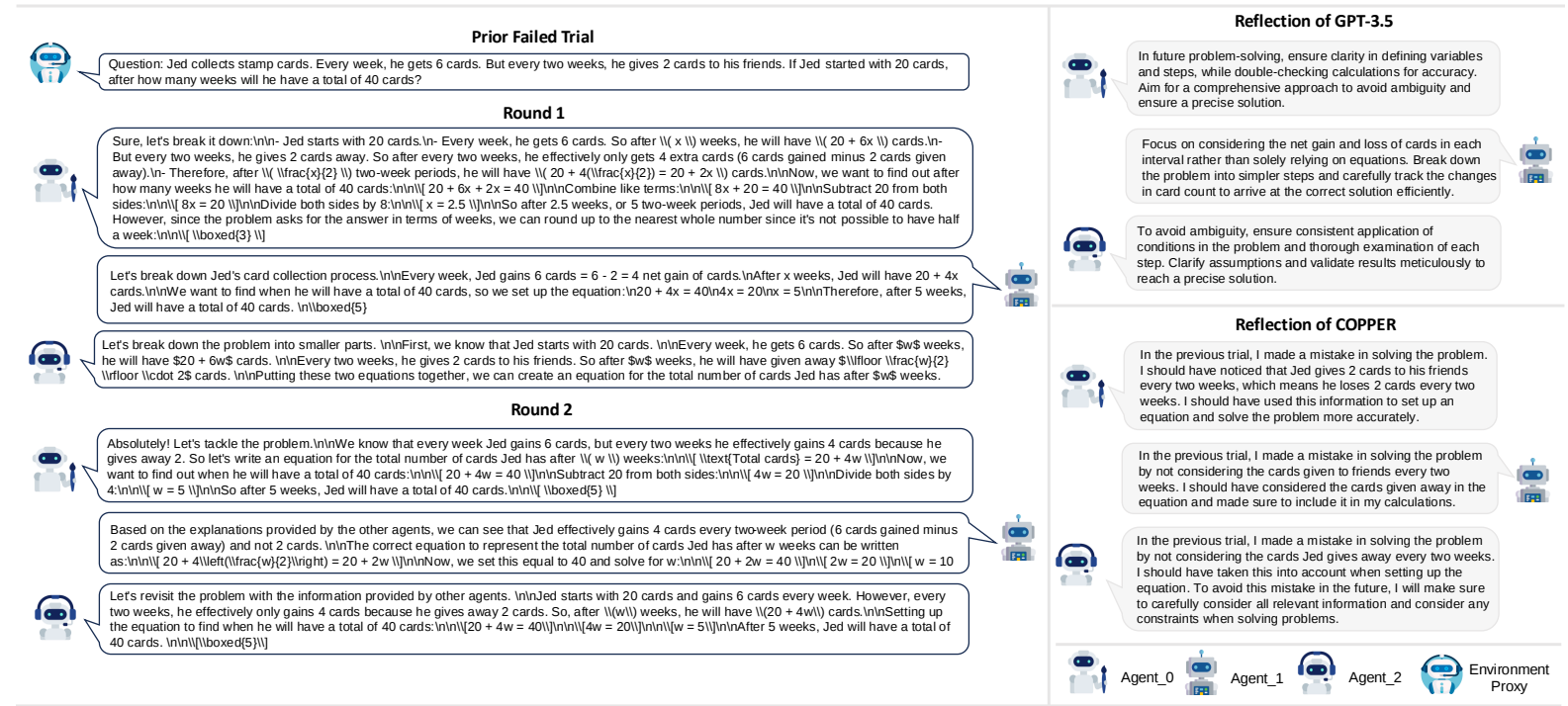


Figure 8: Case study on GSM8K dataset.

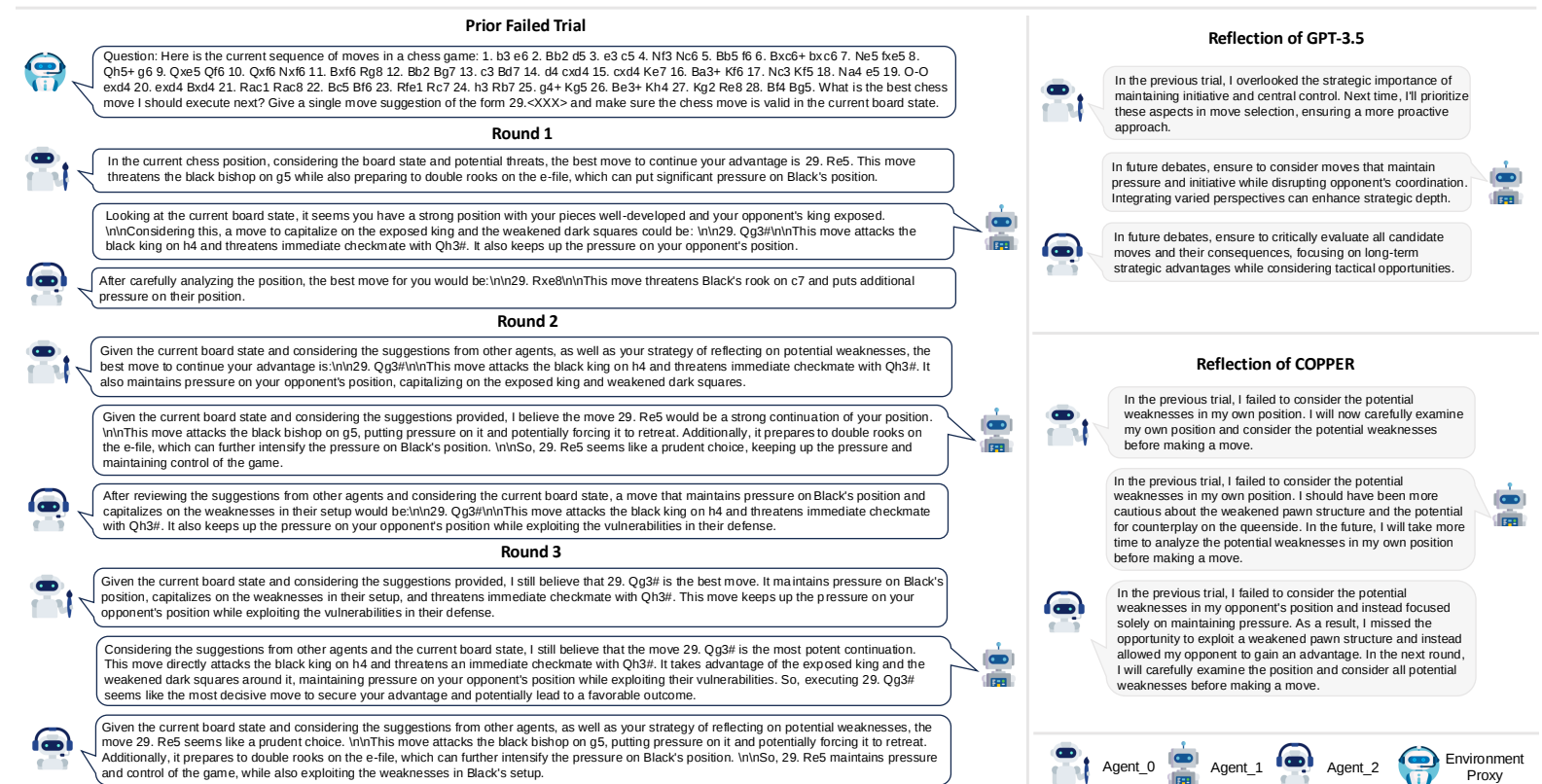


Figure 9: Case study on Checkmate in One Move dataset.