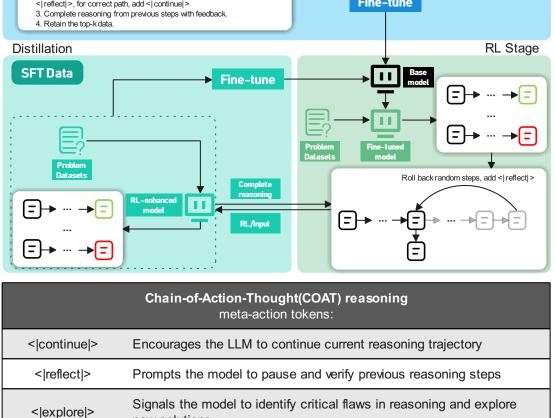
Format Fine-Tuning Data Loop **Iterative** Data Flow Questions Synthesize Wrong step Identification **Sythetic Data** Correct answer LLM-Agent 2 LLM-Agent wrong step 1. Initialize reasoning step trajectories. 2. Give feedback on the chosen step. For wrong steps, add < explore >, Fine-tune <|reflect|>, for correct path, add <| continue|> 3. Complete reasoning from previous steps with feedback 4. Retain the top-k data. **RL Stage** Distillation **SFT Data** Fine-tune



new solutions.