Step 1: Reciprocal Question Answering trains a dual-objective LM for question decomposition and answering response few-shot FIM transformation decomposition **Dataset Transformation SFT** Question Deco-Subquestion mposition (QD) Answering (QA) **Reciprocal Question-Answering Model**

Step 2: Iterative Question Answering Refining

Automated Procedural Supervision: trains a stepwise result verifier Collect LLMs feedback via forward-pass SFT model groundtruth step 2 step n step 1 **Sub-solution Labelling**

Stepwise Verifier

RL fine-tuning:
optimizes against the stepwise verifier

Problem decomposition

SFT model

 \widehat{q}_1

 \hat{q}_2

Context-Guided Decoding

RL Expert Iteration

query

Select high

reward candidates