Stage 1: Data Construction



Two sets of **Human-written Criteria**



Image-Question Pair

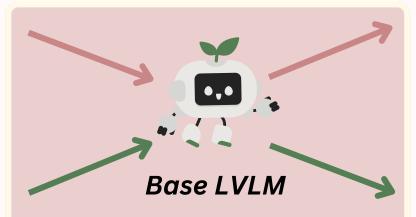
"You should generate a response that follows {Criteria for preferred responses}..."

"You should generate a response that follows {Criteria for rejected responses}..."



Image-Question Pair

Self-Imagination



Preferred Response



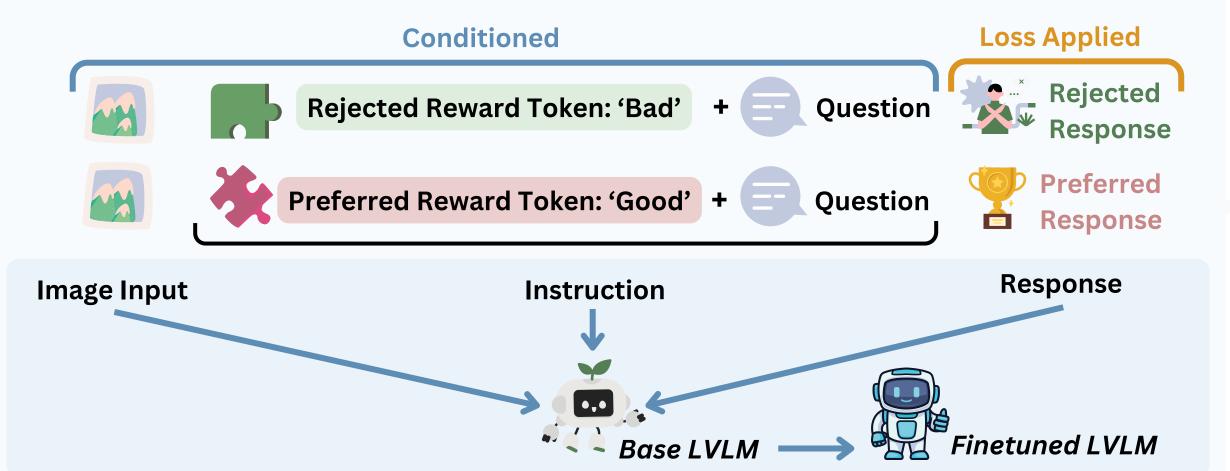
Rejected Response



Stage 2: Conditional Reinforcement Learning

Self Imagination

Prompts



*Inference and Evaluation



+ Question

Preferred Reward Token: 'Good'





Image Input

