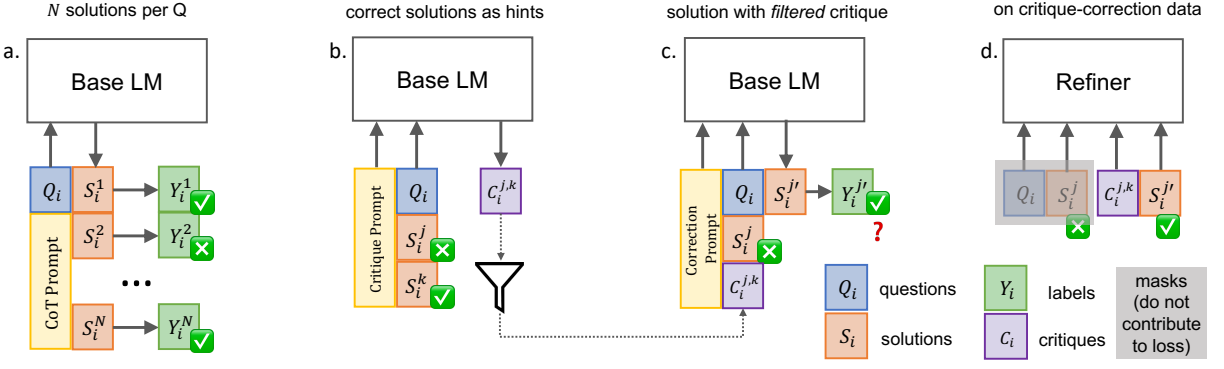

N solutions per Q
correct solutions as hints
solution with filtered critique
on critique-correction data
a.
Base LM
CoT Prompt
b.
Base LM
Critique Prompt
c.
Base LM
Correction Prompt
d.
Refiner
questions
solutions
labels
critiques
masks (do not contribute to loss)