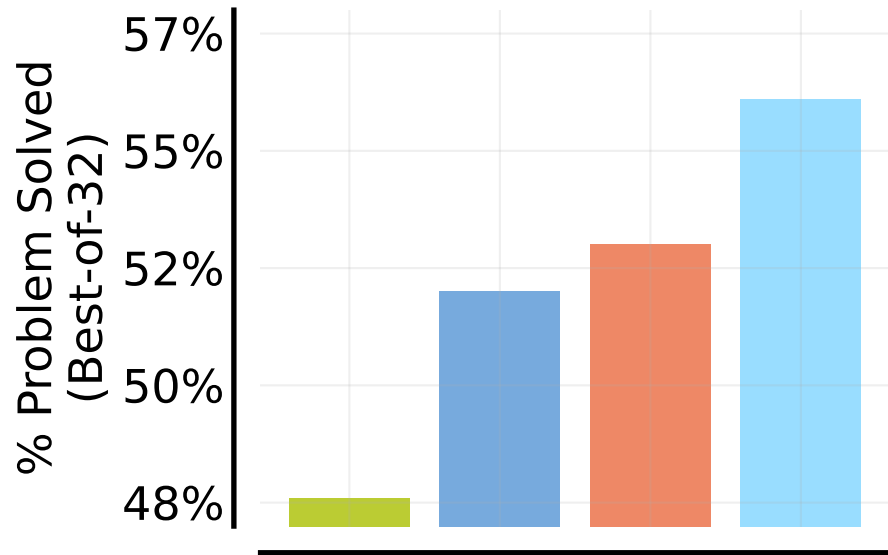




MMLU college_mathematics
 Δ : 47.6% $\rightarrow$ 56.1%



Inference compute scaling
for verification

