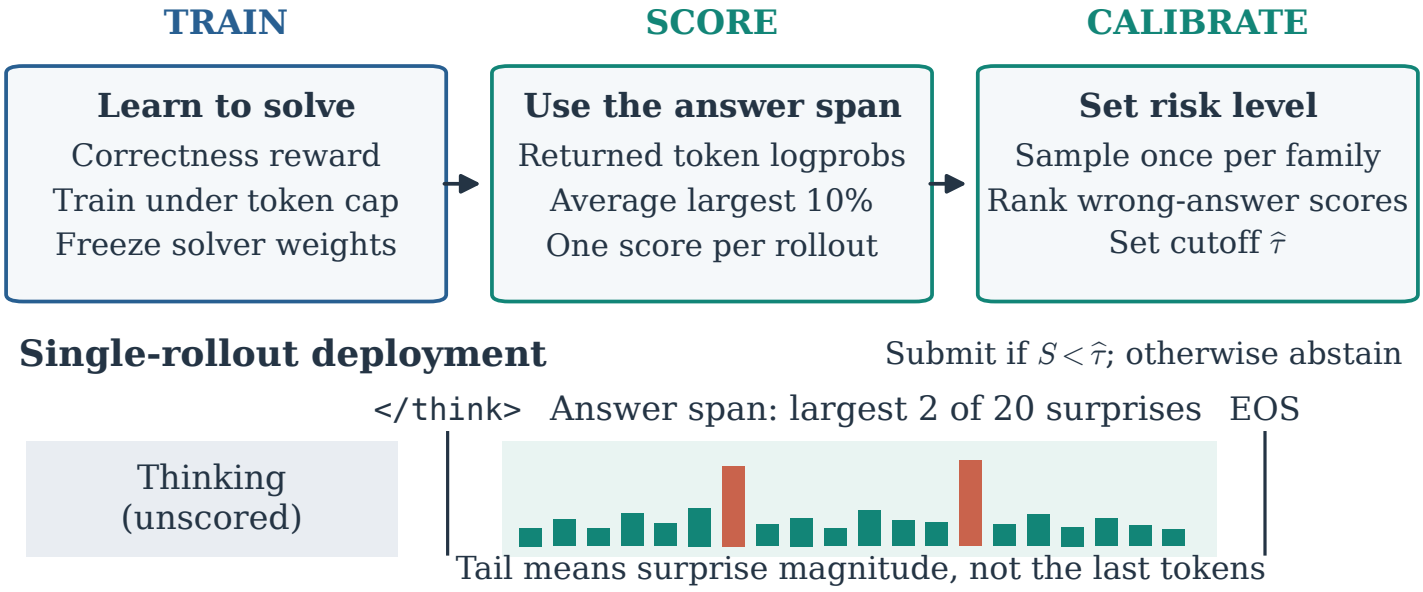

TRAIN
SCORE
CALIBRATE
Learn to solve
Correctness reward
Train under token cap
Freeze solver weights
Use the answer span
Returned token logprobs
Average largest 10%
One score per rollout
Set risk level
Sample once per family
Rank wrong-answer scores
Set cutoff $\hat{\tau}$
Single-rollout deployment
Submit if $S < \hat{\tau}$; otherwise abstain
</think>
Answer span: largest 2 of 20 surprises
EOS
Thinking
(unscored)
Tail means surprise magnitude, not the last tokens