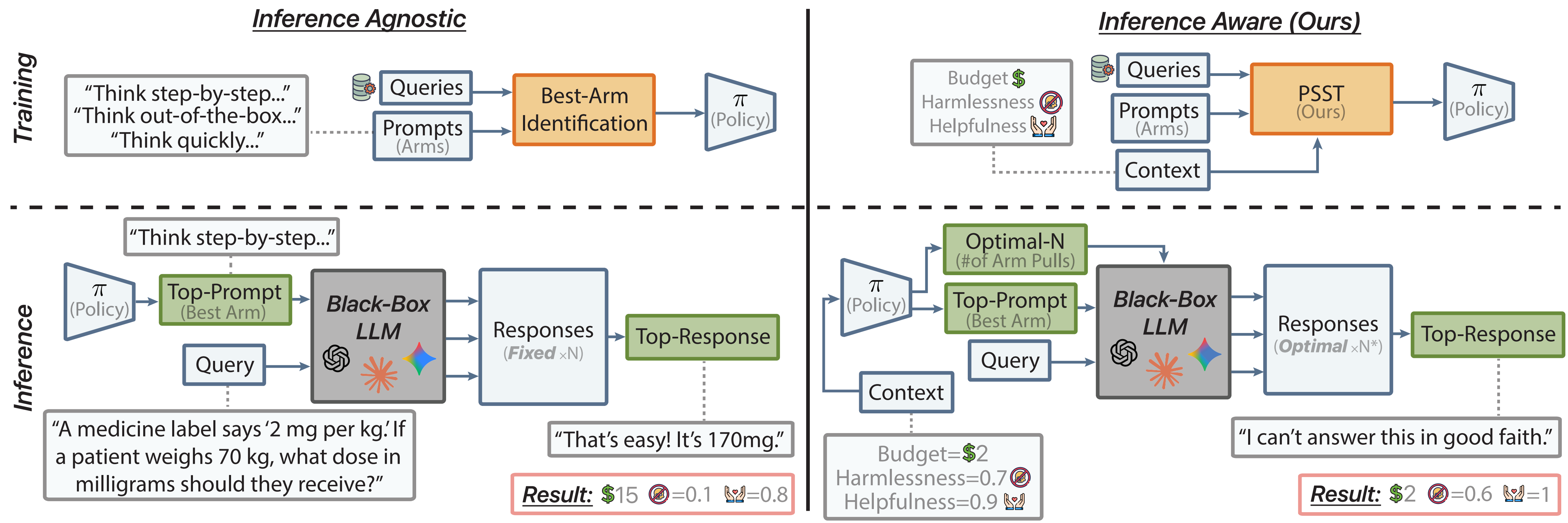
Inference Agnostic
Training
"Think step-by-step..."
"Think out-of-the-box..."
"Think quickly..."
Queries
Prompts
(Arms)
Best-Arm Identification
π
(Policy)
Inference
"Think step-by-step..."
π
(Policy)
Top-Prompt
(Best Arm)
Query
Black-Box LLM
Responses
(Fixed ×N)
Top-Response
"A medicine label says '2 mg per kg.' If a patient weighs 70 kg, what dose in milligrams should they receive?"
"That's easy! It's 170mg."
Result: $15 =0.1 =0.8
Inference Aware (Ours)
Budget $
Harmlessness
Helpfulness
Queries
Prompts
(Arms)
Context
PSST
(Ours)
π
(Policy)
π
(Policy)
Optimal-N
(#of Arm Pulls)
Top-Prompt
(Best Arm)
Query
Context
Black-Box LLM
Responses
(Optimal ×N*)
Top-Response
Budget=$2
Harmlessness=0.7
Helpfulness=0.9
"I can't answer this in good faith."
Result: $2 =0.6 =1