



Human




Data



LLM



Question  → Which model will work best?



Training Set Models: 

Selected Model: 



(a) Router: Memorizing Models



Question  → How will the model perform?



Training Set Models: 

Unseen Models: 

Analysis: Based on the model's past behaviors...

Predicted Performance: {len: 600, correct: no}



Budget

Low  High

Selected Model

Cheap  Expensive



(b) SCOPE: Reasoning on Model Behaviors