



Motivation : Search is not always helpful.

A Search helps on long-tail factual questions

What is Lothrop Stoddard's occupation?

Without Search

Relies on
parametric memory

✗ Wrong Answer

With Search

Retrieved evidence identifies:
writer / journalist

✓ Correct Answer

B Search can mislead on ambiguous questions

What is the capital of Georgia?

Without Search

Detects ambiguity
"Georgia the country or the state?"

✓ Correct Answer

With Search

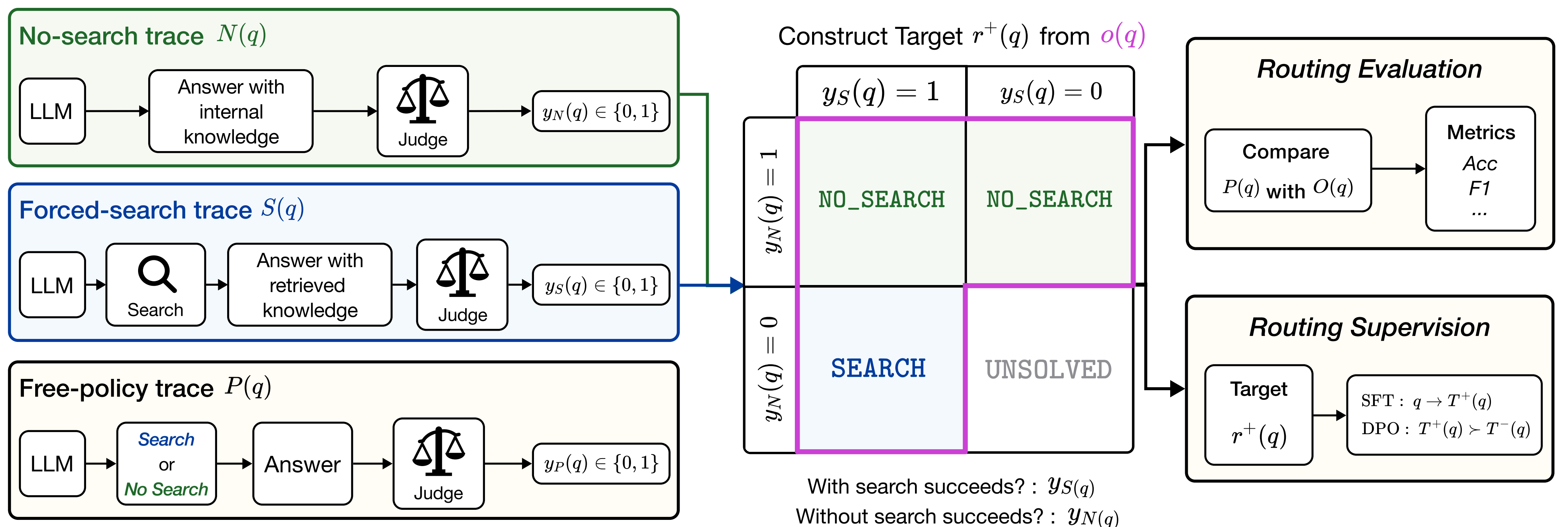
Finds one interpretation
Tbilisi or Atlanta

✗ Premature Answer

Whether to search depends on what the model knows and what the question requires.
Over-searching wastes cost and risks misleading evidence. under-searching misses recoverable facts.

Method: Counterfactual Routing Oracle

We compare task success with and without search for the same question to infer whether search is needed.



Results: Counterfactual Supervision Improves Search Routing

Model	Policy	Macro-F1	Routing Acc.	Cond. Over-search	Cond. Under-search	FA Acc.	Amb. Acc.
Gemma E2B	P_{pre}	0.7082	0.7574	0.5425	0.0598	0.6897	0.7170
Gemma E2B	SFT	0.8207	0.8292	0.1928	0.1574	0.7759	0.8868
Gemma E2B	Preference Opt.	0.8235	0.8329	0.2059	0.1434	0.8103	0.9434
Qwen3.5-4B	P_{pre}	0.7053	0.7148	0.3649	0.2272	0.8732	0.6531
Qwen3.5-4B	SFT	0.8207	0.8263	0.2284	0.1339	0.9296	0.8776
Qwen3.5-4B	Preference Opt.	0.8365	0.8427	0.2312	0.1034	0.9296	0.9184

Takeaways

- SFT improves search-routing Macro-F1 for both models.
- Preference Optimization further refines the search/no-search boundary.
- Improvements differ by model: Gemma reduces over-search, while Qwen reduces under-search.
- Gains come from correcting instance-level routing, not simply changing the search rate.