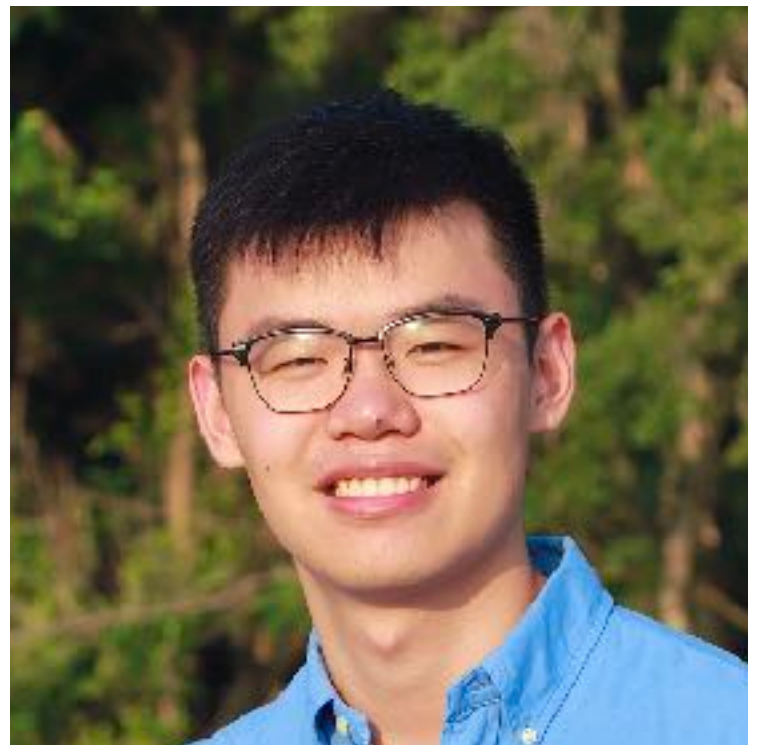
Large Language Models as Commonsense Knowledge for Large-Scale Task Planning



Zirui Zhao

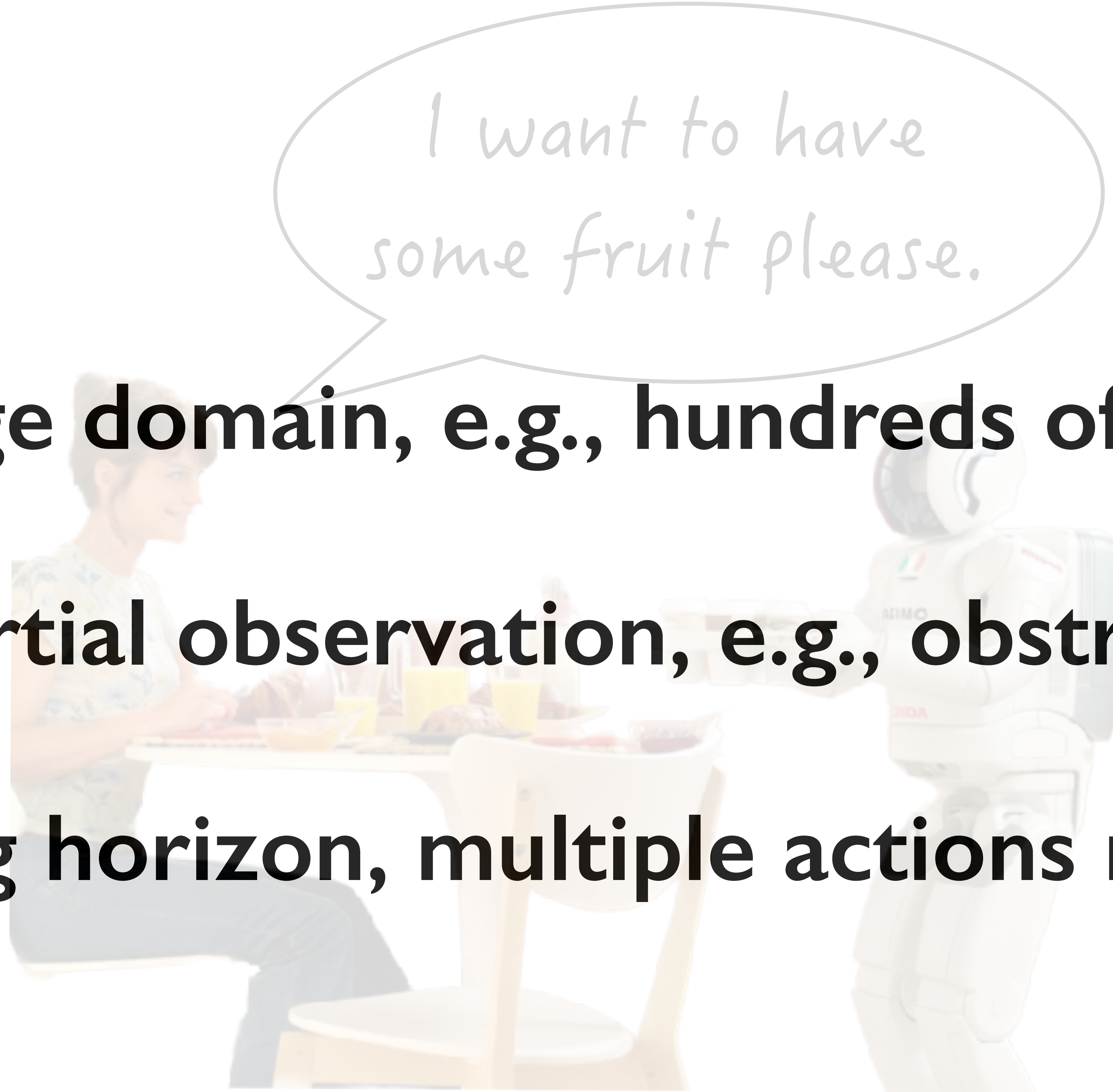
Wee Sun Lee

David Hsu

NeurIPS 2023

NUS | Computing
National University of Singapore

Planning in large-scale environments



Large domain, e.g., hundreds of objects

Partial observation, e.g., obstruction

Long horizon, multiple actions required

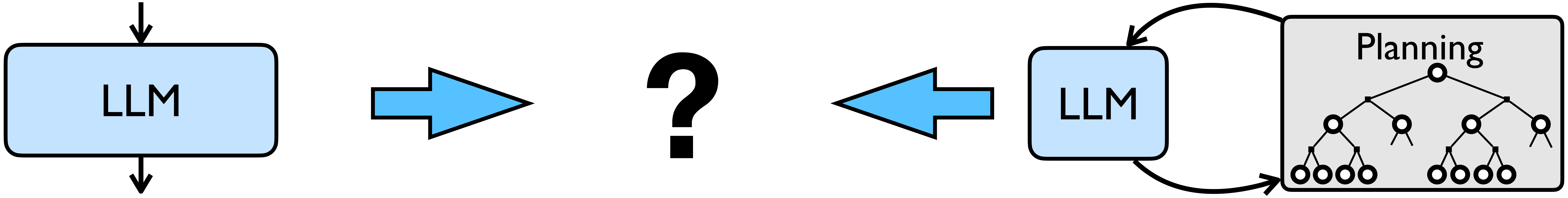
Planning in large-scale environments

How to solve the challenging large-scale planning problems?

Planning with large language models

I want to have some fruit please

Action: Go to kitchen



1. Go to kitchen; 2. Open fridge;
3 …

Observation: fridge in the kitchen;
Next state: you at kitchen, …;
Reward: …

LLM as a policy

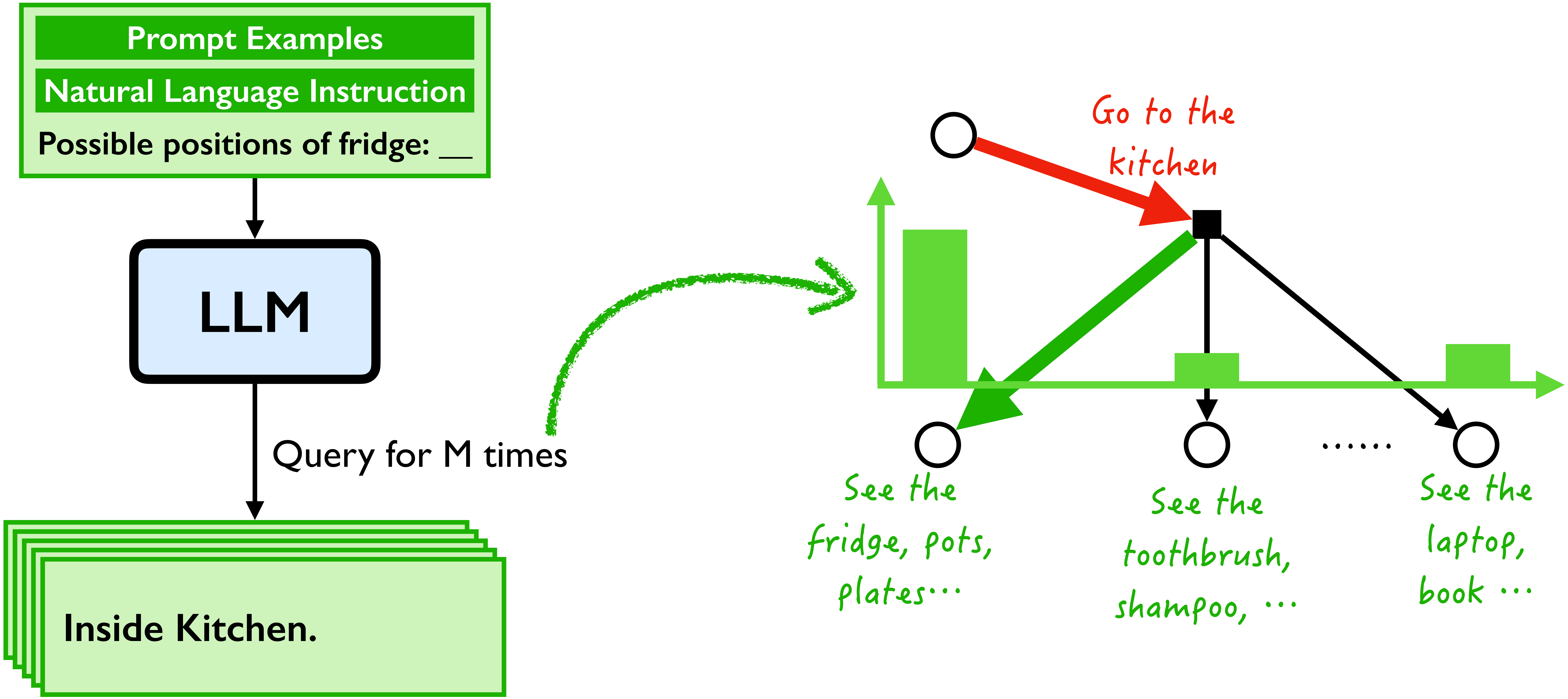
LLM as a world model

E.g., SayCan; Inner Monologue;
Voyager; …

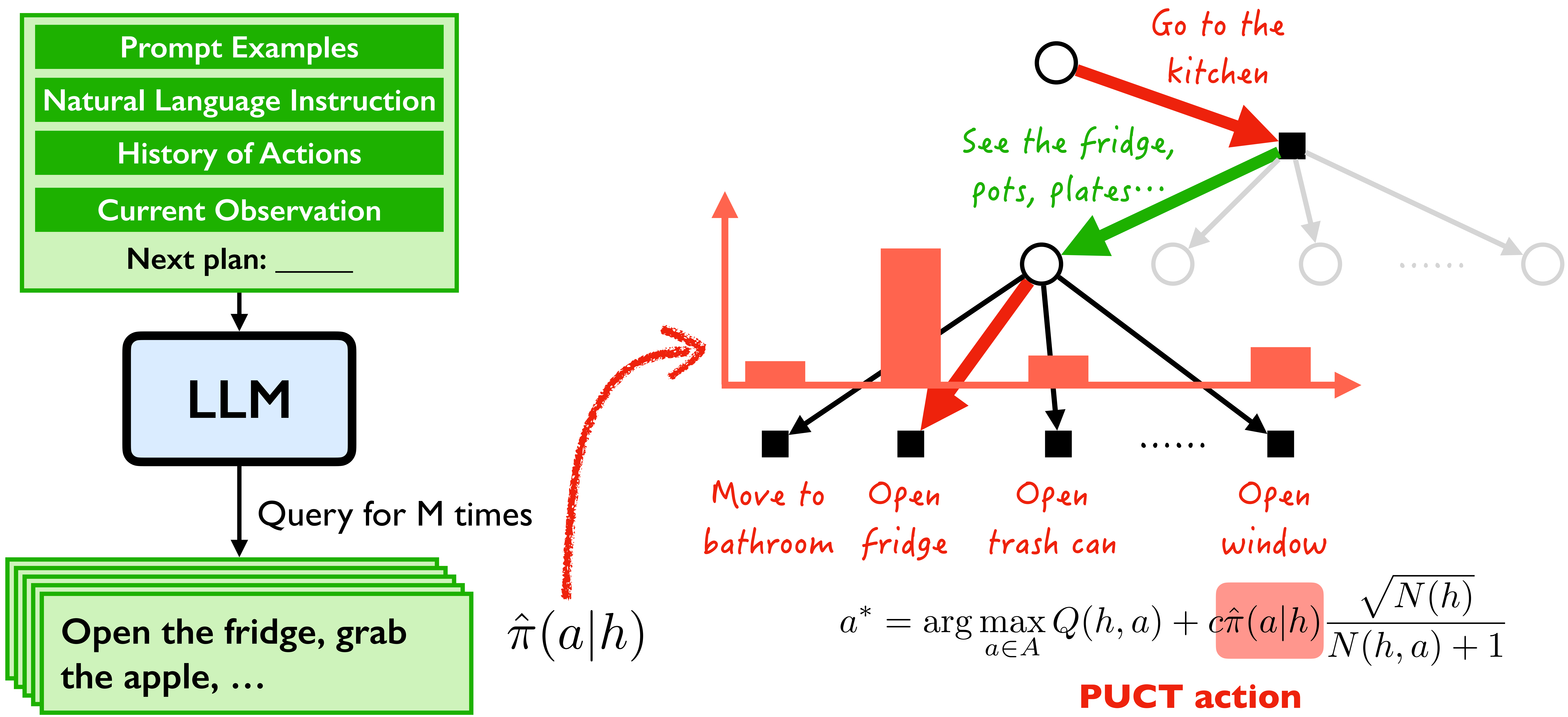
LLM as a ***world model*** + LLM as a ***policy***

- LLM as *world model* and *policy* in *planning algorithm* (Monte Carlo Tree Search)
  - LLM world model improves the LLM policy's accuracy
  - LLM policy as a search heuristic to help the planning

LLMs as Commonsense World Model



LLMs as Commonsense Heuristic Policy

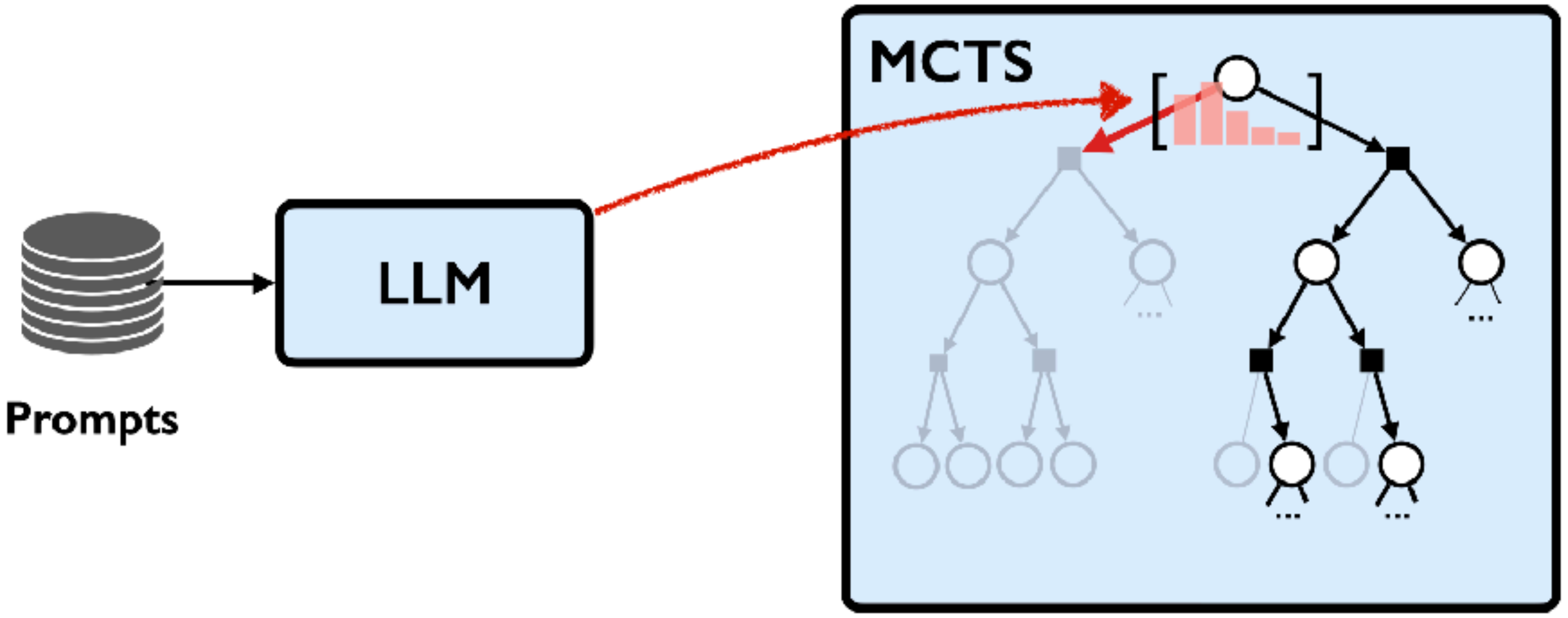


$$a^* = \arg\max_{a \in A} Q(h,a) + c\hat{\pi}(a|h)\frac{\sqrt{N(h)}}{N(h,a)+1}$$

PUCT action selection

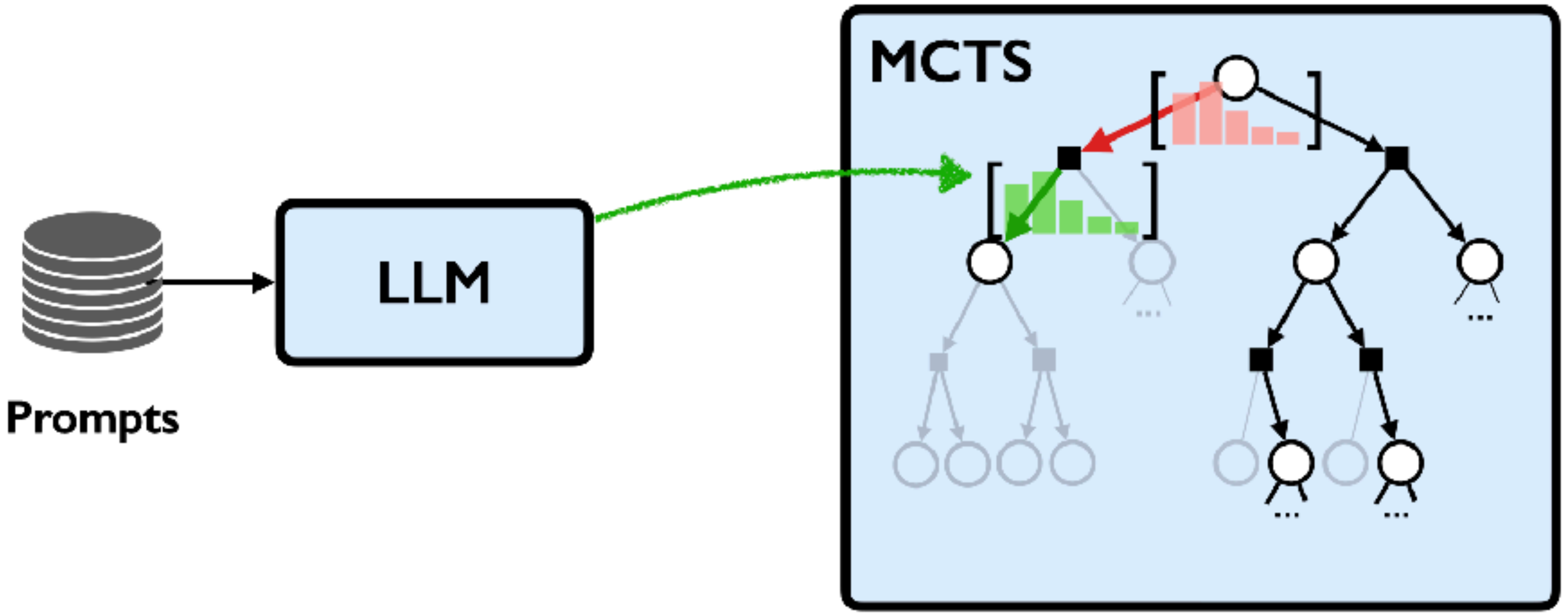
Monte Carlo Planning with LLM

- Sampling from belief tree for approximate planning
    - Action selection: select action biasedly according to commonsense



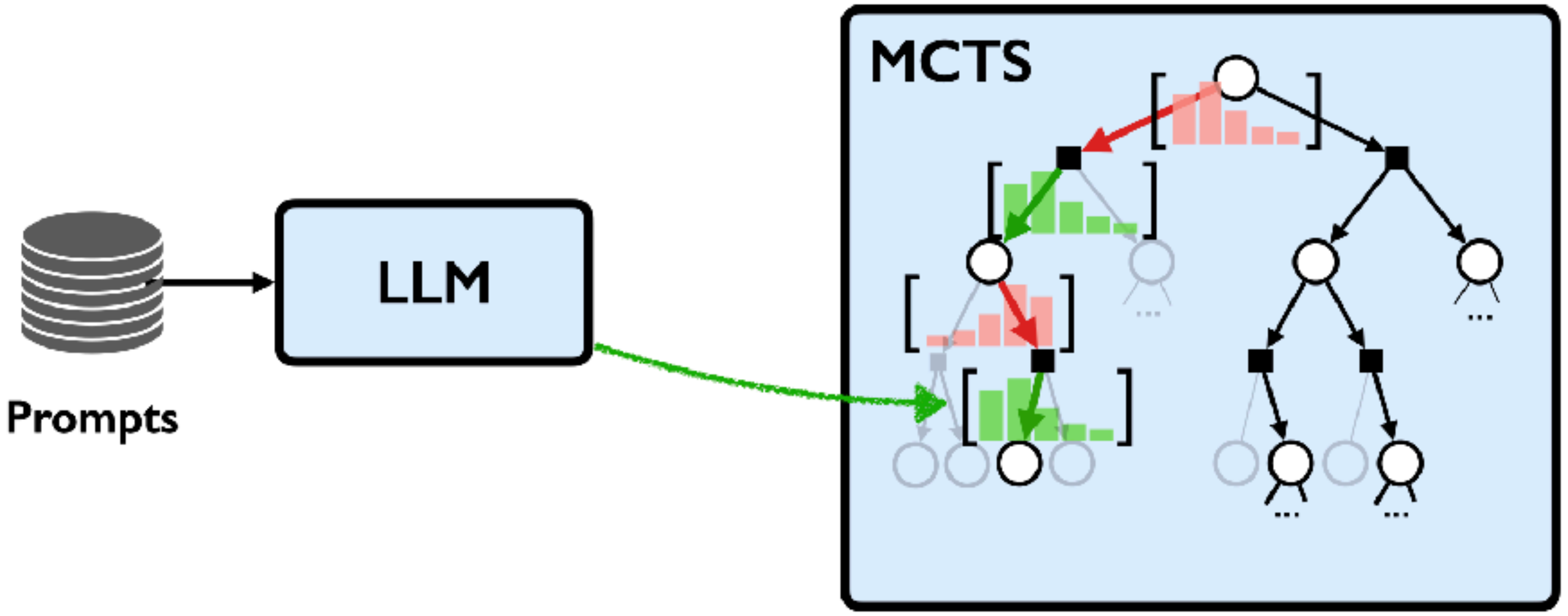
Monte Carlo Planning with LLM

- Sampling from belief tree for approximate planning
  - Action selection: select action biasedly according to commonsense
  - Observation sampling: sample observation according to commonsense



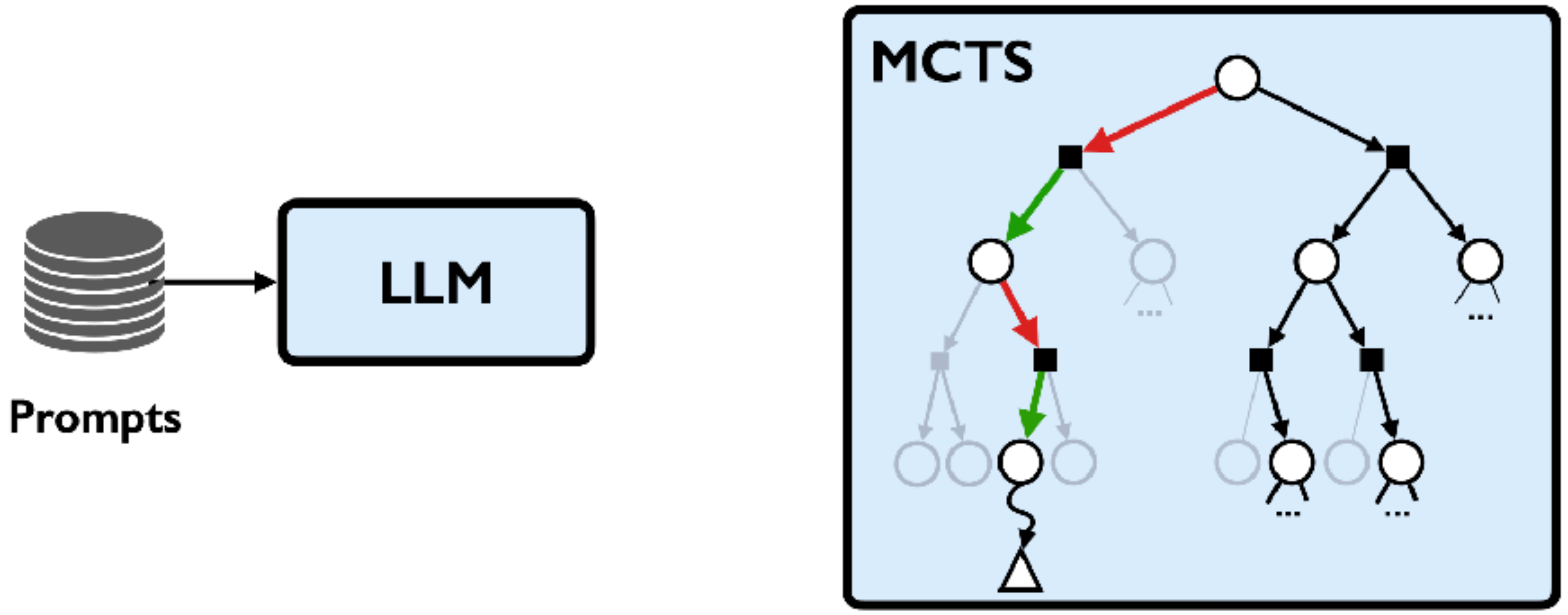
Monte Carlo Planning with LLM

- Sampling from belief tree for approximate planning
  - Action selection: select action biasedly according to commonsense
  - Observation sampling: sample observation according to commonsense



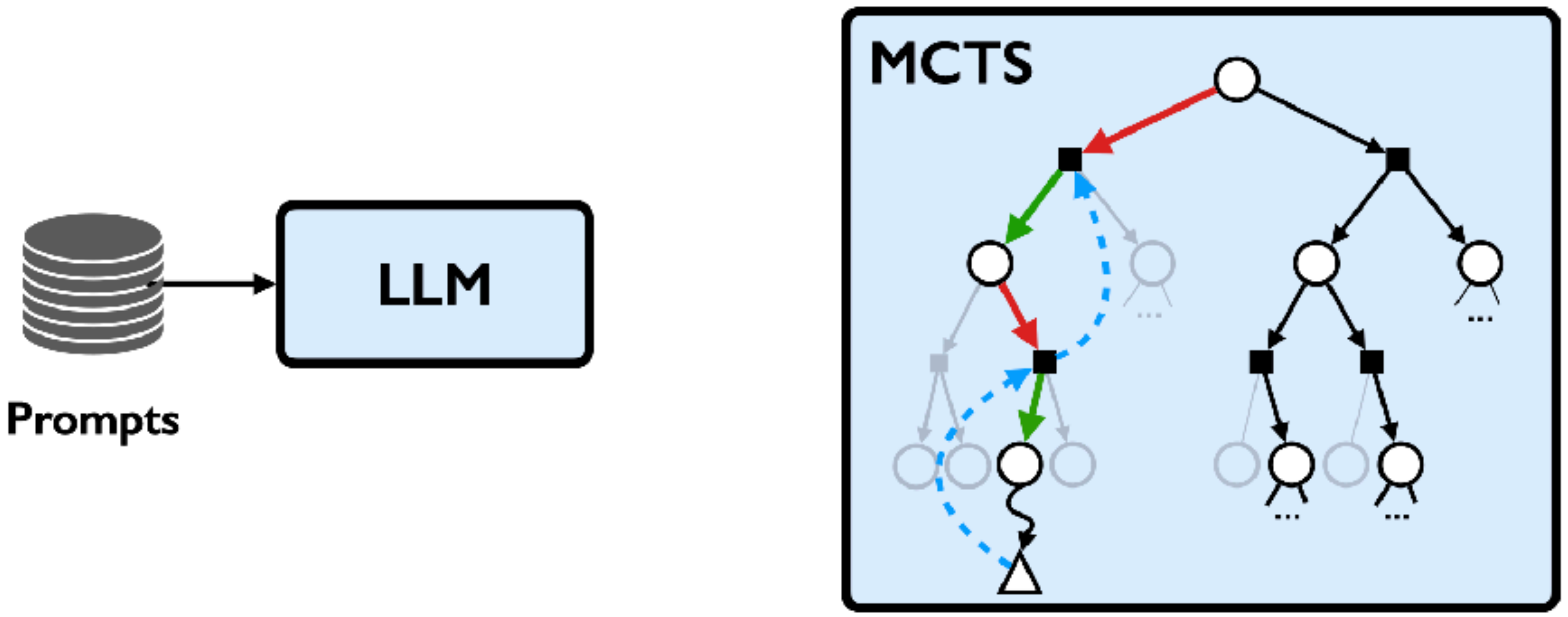
Monte Carlo Planning with LLM

- Sampling from belief tree for approximate planning
  - Action selection: select action biasedly according to commonsense
  - Observation sampling: sample observation according to commonsense
  - Expansion & Rollout: get the reward



Monte Carlo Planning with LLM

- Sampling from belief tree for approximate planning
  - Action selection: select action biasedly according to commonsense
  - Observation sampling: sample observation according to commonsense
  - Expansion & Rollout: get the reward
  - Backup: update the estimated Q function

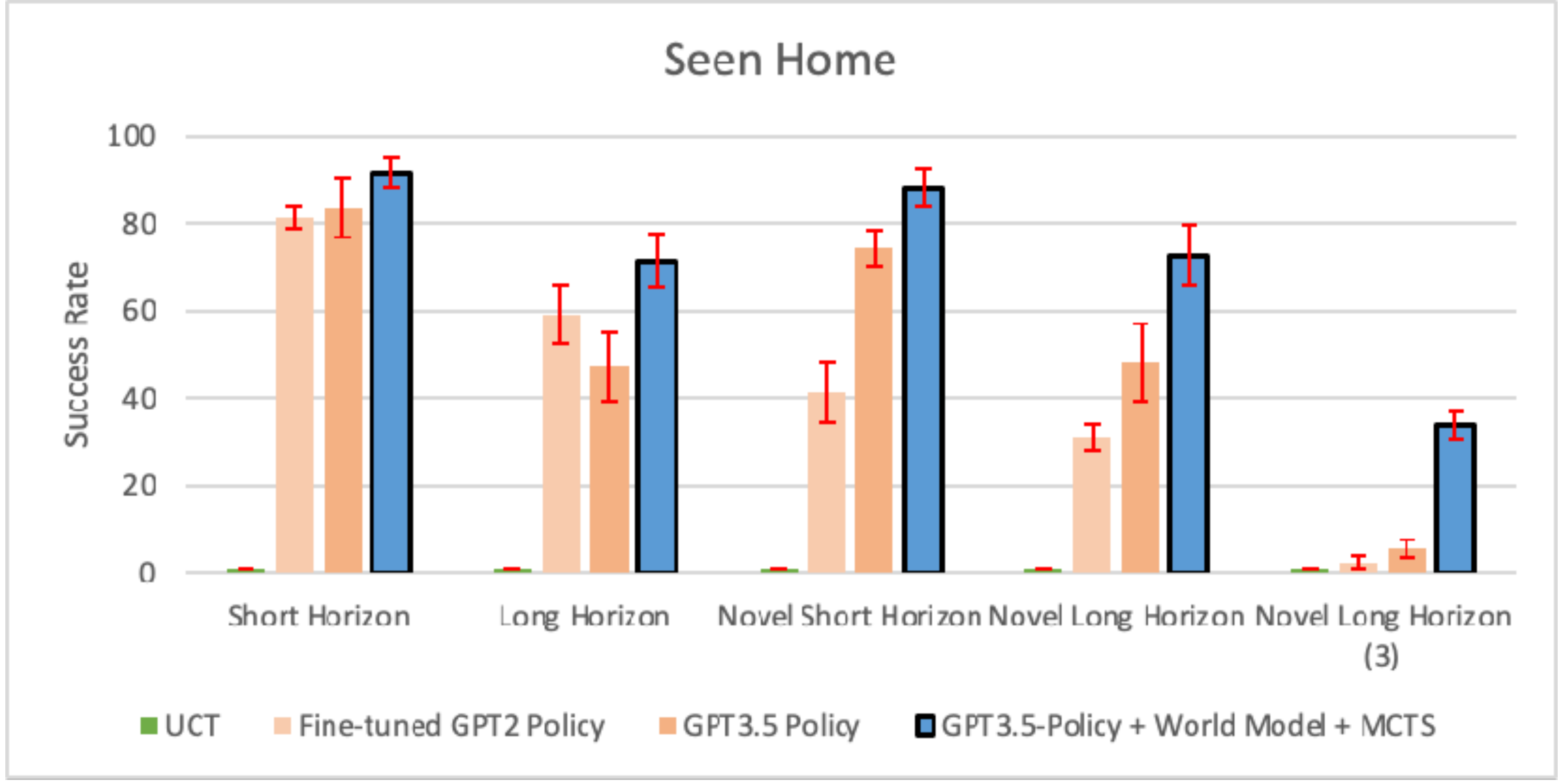


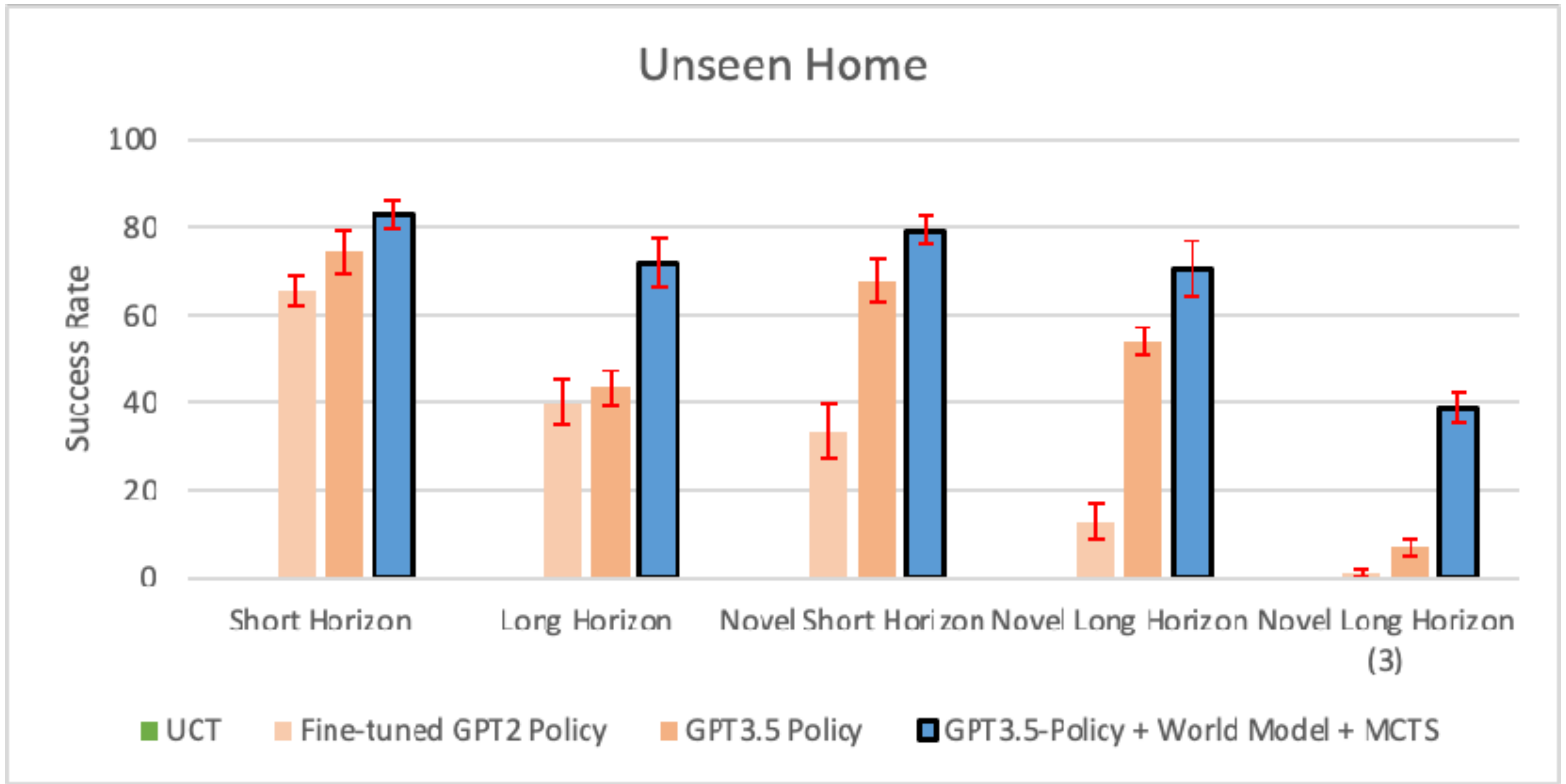
Experiments

- VirtualHome simulator
- Task: object rearrangements in household environments
  - Simple v.s. compositional tasks
  - In-distribution v.s. novel tasks
- Baselines:
  - LLM as world model: Upper confidence tree (UCT) without heuristic
  - LLM as Policy: GPT3.5 and GPT2 policy

Experimental results

- LLM as both the world model and policy outperforms either alone
  - A more accurate LLM world model improve the accuracy of LLM policy
  - LLM policy guides planning to make it more efficient

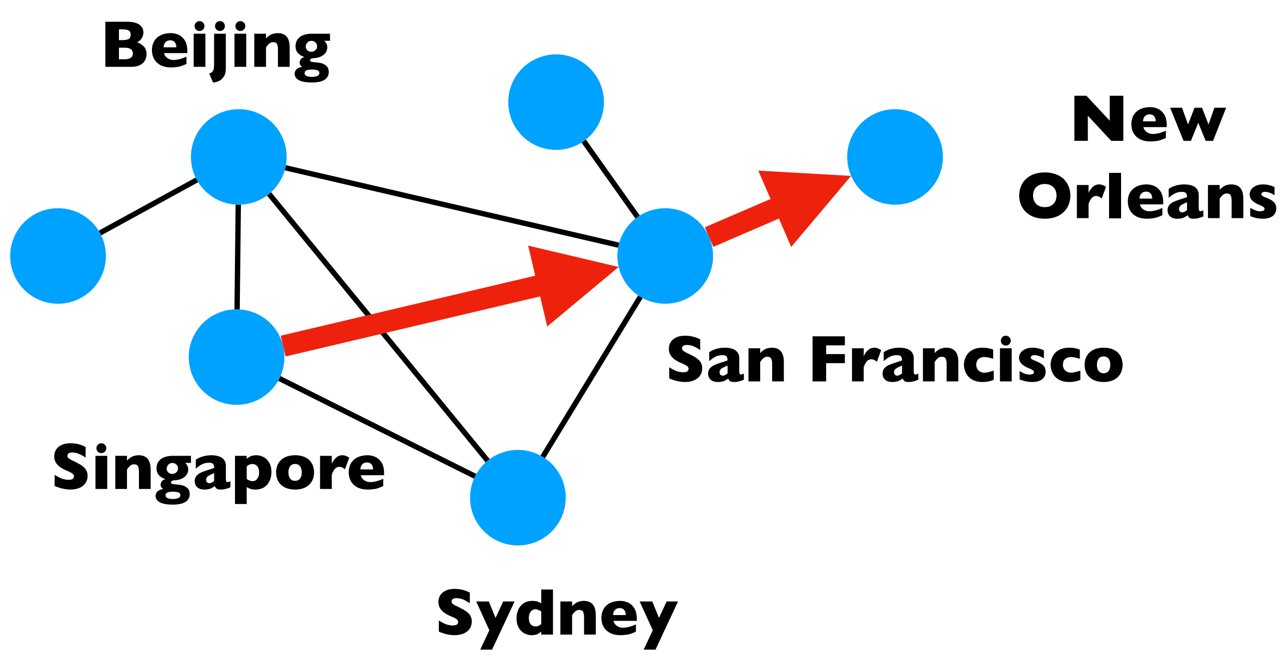


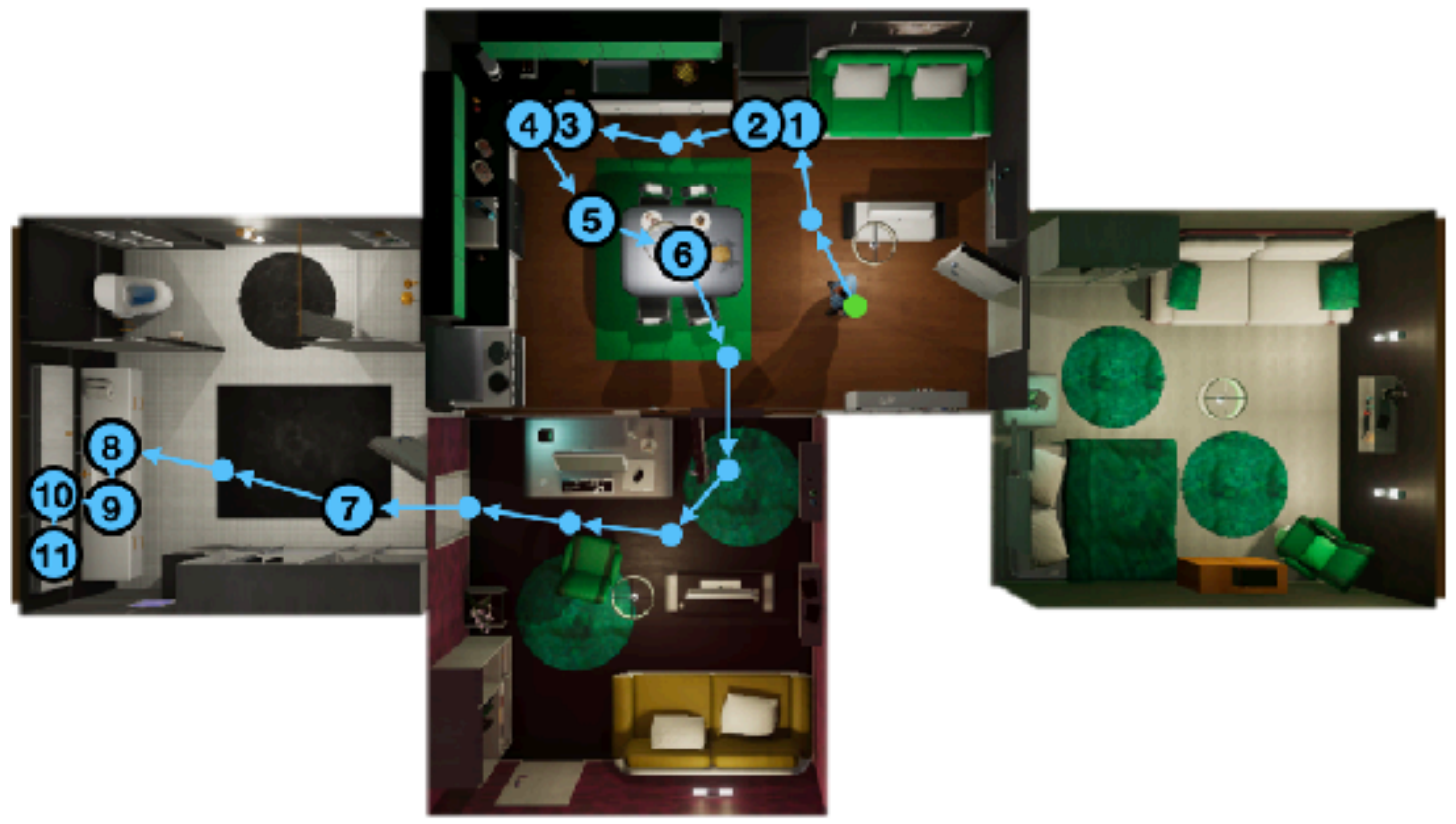


LLM as world model or policy?

- Using LLM as ***World Model*** or ***Policy***, which is better?
- *Minimum Description Length* (MDL): method with ***shorter description length*** has smaller generalization error[1]
- Analysis and experiments: multi-digit multiplication, travel planning, object rearrangement, …

14128 x 8634 = ?





Instruction: Put one apple on the kitchen table and one toothbrush inside the bathroom cabinet.
1: Walk to fridge
2: Open fridge
3: Walk to apple
4: Grab apple
5: Walk to kitchen table
6: Put apple on kitchen table
7: Walk to bathroom
8: Walk to toothbrush
9: Grab toothbrush
10: Open bathroom cabinet
11: Put toothbrush inside bathroom cabinet

[1] Shai Shalev-Shwartz and Shai Ben-David. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.

Summary

- LLM as world model and policy outperforms either one
- Choose between LLM world model and policy? Use MDL principle: shorter description length is better

Thank You!

Paper and Code

Contact
ziruiz@comp.nus.edu.sg

Website
https://llm-mcts.github.io

Example: multi-digit multiplication

- LLM Policy
  - Table with inputs and results
  - Description length: $O(n10^n)$
- LLM Model + algorithm
  - Single-digit multiplication table by LLM
  - Algorithm
  - Description length: constant
- Empirical results

LLM Policy

	0	1	2	...	10^n-1
0	0	0	0	...	0
1	0	1	2	...	10^n-1
2	0	2	4		...
...	...	...		...	...
10^n-1	0	10^n-1	...	...	...

LLM World Model + Algorithm

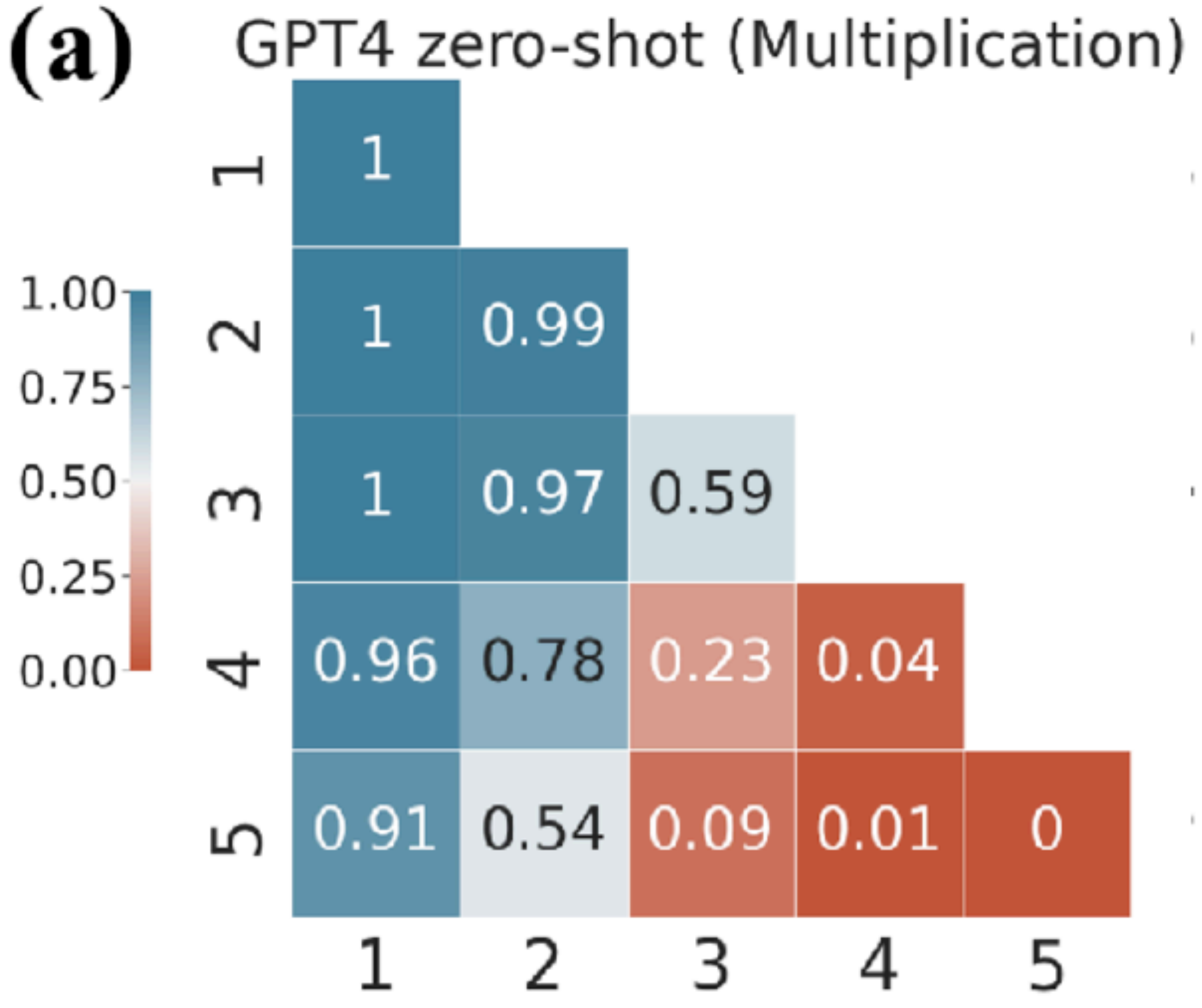
	0	1	...	9
0	0	0	...	0
1	0	1	...	9
...	...	...	...	...
9	0	9		81

```
function multiply (x[1..p], y[1..q]):
  // multiply x for each y[i]
  for i = q to 1
    carry = 0
    for j = p to 1
      t = x[j] * y[i]
      t += carry
      carry = t // 10
      digits[j] = t mod 10
    summands[i] = digits

  // add partial results (computation not shown)
  product = ∑_{i=1}^{q} summands[q+1-i] · 10^{i-1}
  return product
```

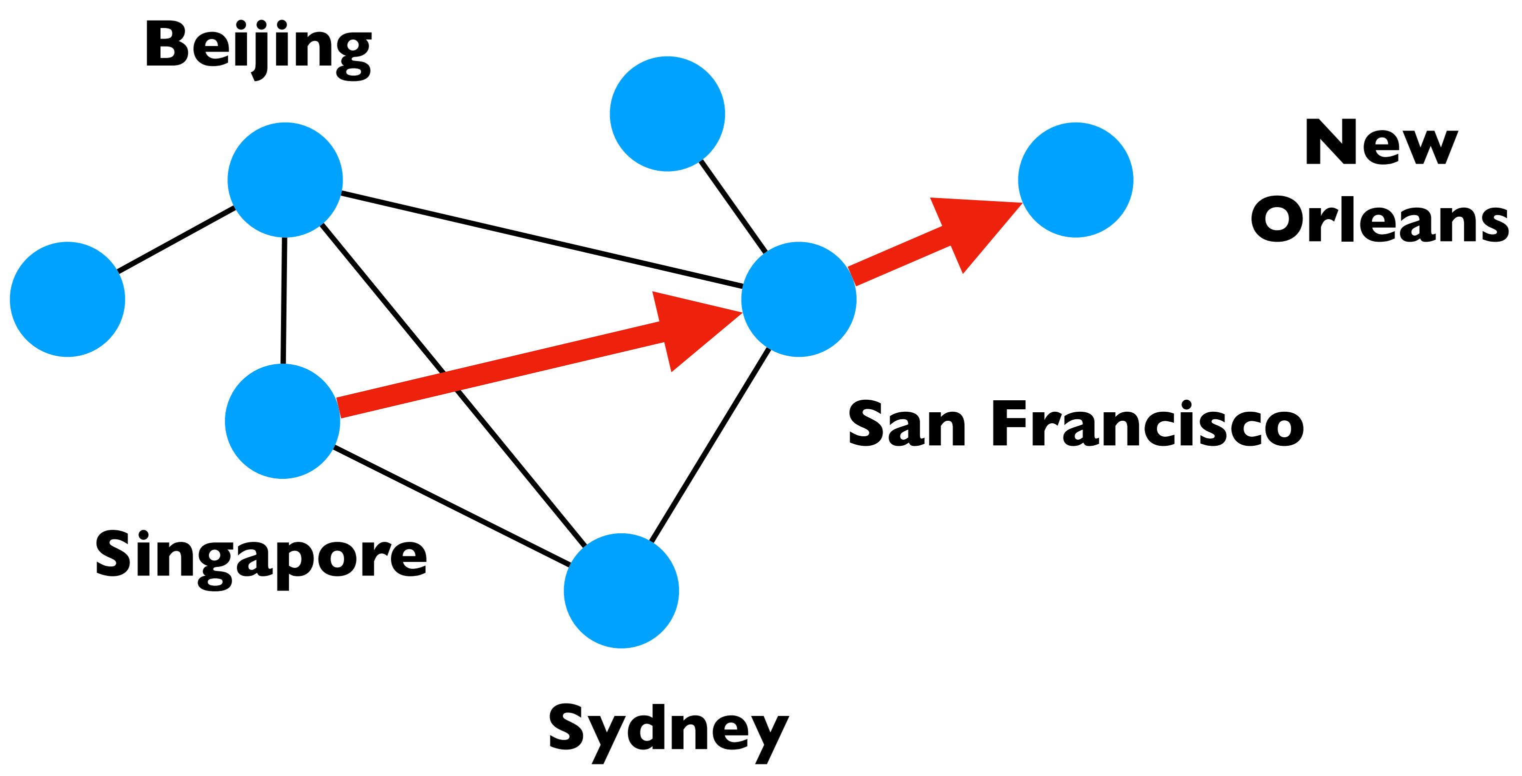
Example: multi-digit multiplication

- LLM Policy
  - Table with inputs and results
  - Description length: $O(n10^n)$
- LLM Model + algorithm
  - Single-digit multiplication table by LLM
  - Algorithm
  - Description length: constant
- Empirical results



Example: travel planning

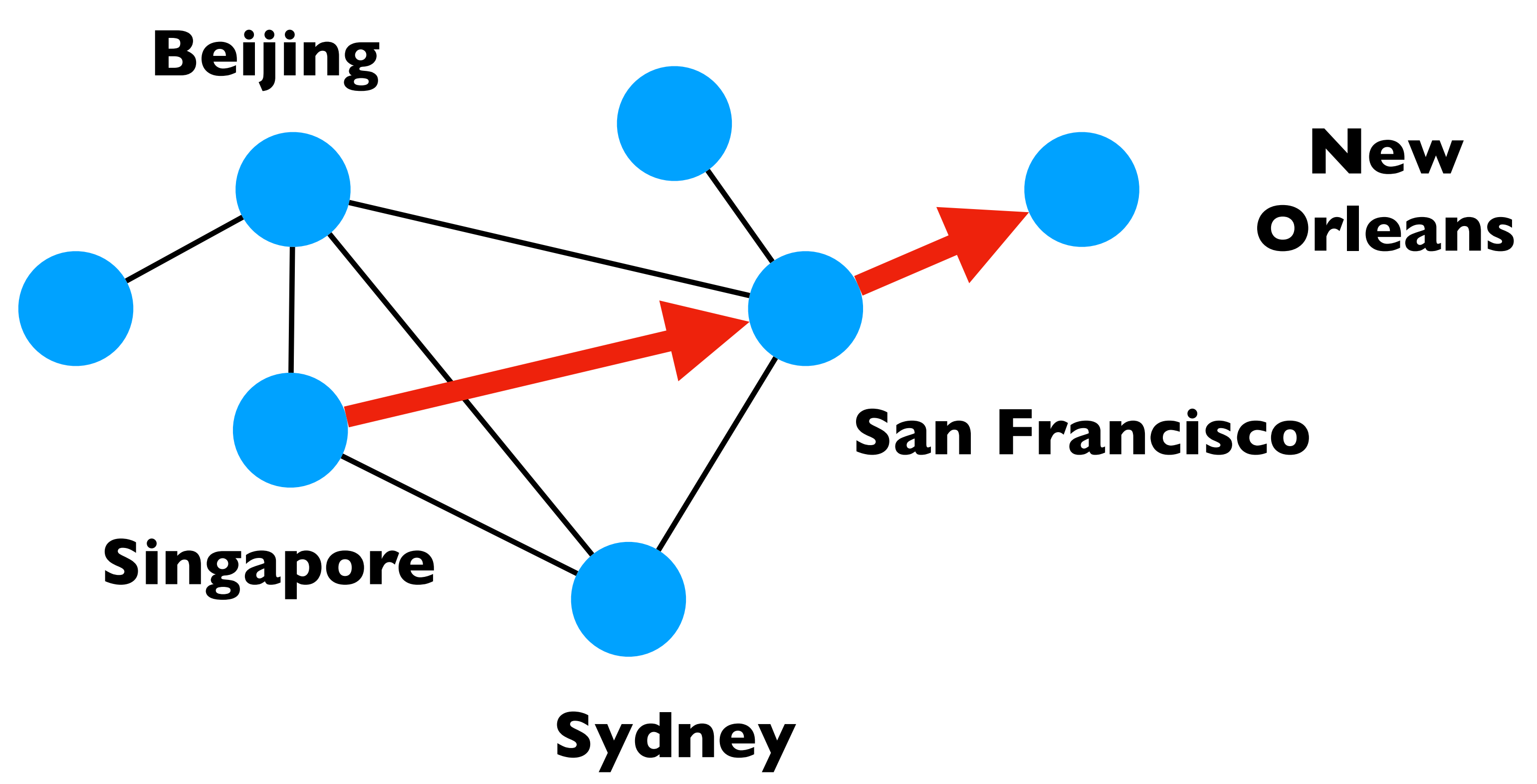
- Problem: predict flight routes between given cities
- LLM Policy: table of travel
  - Description length: $O(n^2 \log n)$
- LLM World Model solution: flight graph+search
  - Description length: $O(n \log n)$
- Results: LLM World Model solution works better

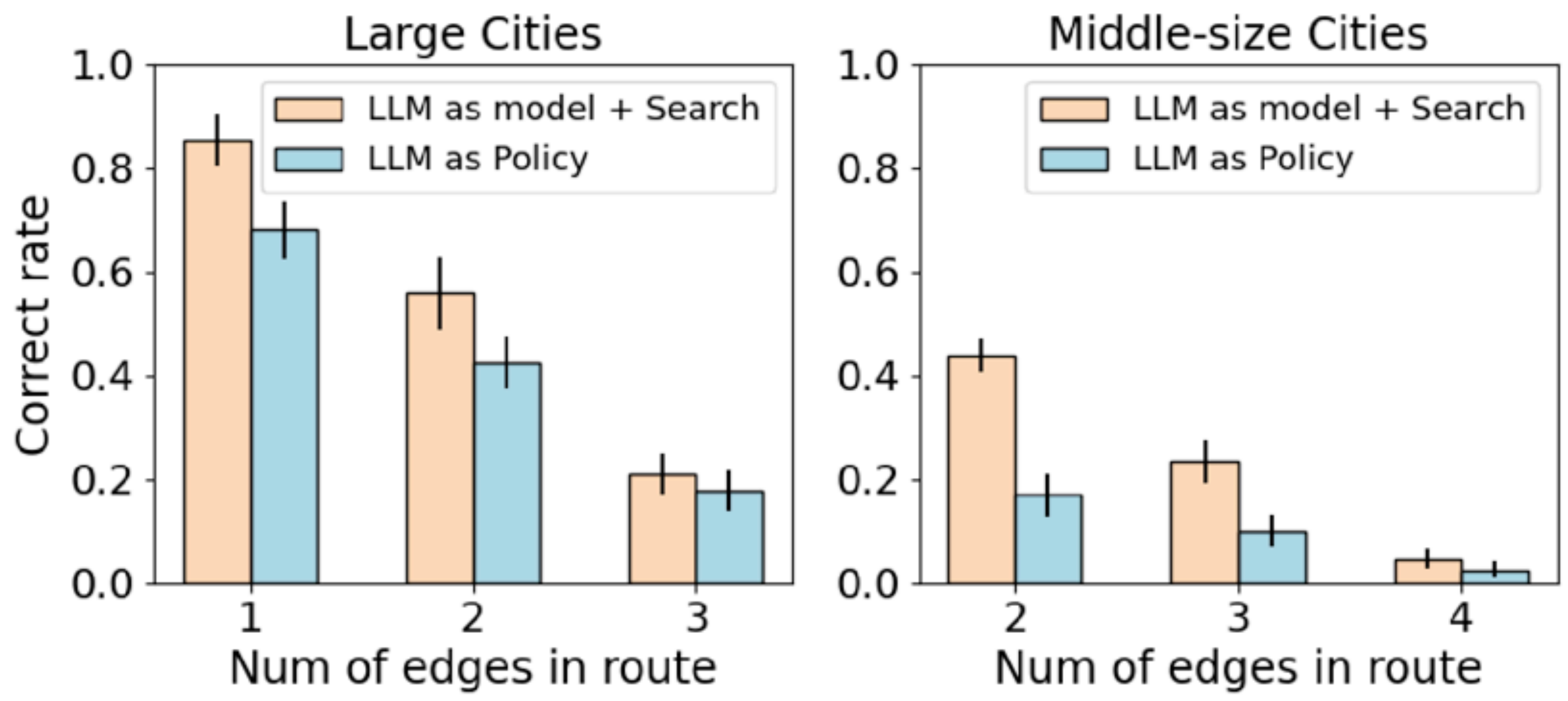


Current\goal	New Orleans	Sydney	...
Singapore	San Francisco	Sydney	
Sydney	San Francisco	—	
San Francisco	New Orleans	Sydney	
...			

Example: travel planning

- Problem: predict flight routes between given cities
- LLM Policy: table of travel
  - Description length: $O(n^2 \log n)$
- LLM World Model solution: flight graph+search
  - Description length: $O(n \log n)$
- Results: LLM World Model solution works better





Example: object rearrangement

- Consider a house with n objects, m containers, and k rooms
- LLM policy description length: $O(mn \log(m+k))$
- LLM world model description length: $O((m+n)\log(m+k))$
- Both: LLM world model + LLM policy heuristic
  - LLM Policy helps search algorithm
  - LLM world model is more accurate and improve LLM policy