



GRIDSPACE IAP 2024 LECTURE 2
TOOLS FOR LLM PLANNING

January 11, 2023

TODAY'S ROADMAP

- Planning
- How do humans do planning
- Search augmented Planning
- LLM as Learner
- LLM as Tool Retriever

PLANNING

Planning

[Article](#) [Talk](#)

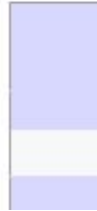
F

From Wikipedia, the free encyclopedia

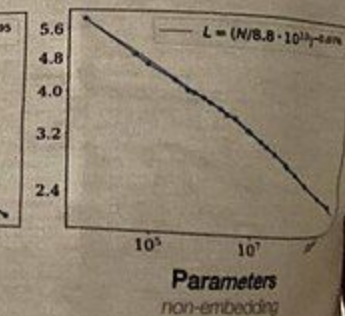
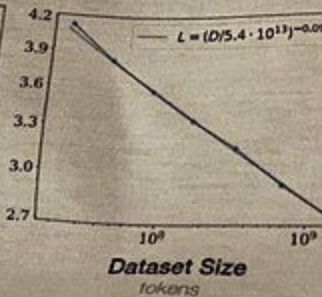
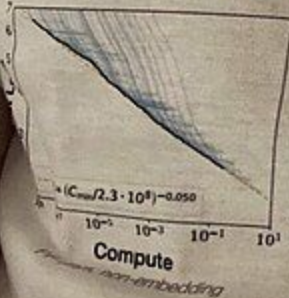
"Forethought" redirects here. For the defunct software company, see [Forethought, Inc.](#)

This article is about the concept. For urban planning, often referred to as simply "planning", see [urban planning](#).

Planning is the [process](#) of [thinking](#) regarding the activities required to achieve a desired [goal](#). Planning is based on foresight, the fundamental capacity for [mental time travel](#). The evolution of forethought, the capacity to think ahead, is considered to have been a prime mover in [human evolution](#).^[1] Planning is a fundamental property of intelligent behavior. It involves the use of logic and imagination to visualise not only a desired result, but the steps necessary to achieve that result.



**SCALE IS ALL YOU NEED -
AGI IS COMING**



Resources » Articles » The Top 7 AI Trip Planner Tools

The Top 7 AI Trip Planner Tools



The Upwork Team

Aug 9, 2023 | 10 Min Read

AI Services

Article

5 AI tools for meal planning and nutrition



Priya Ranjani Mohan

🌟 LinkedIn Top Voices - Machine Learning 🌟 | Management Consultant
at KPMG | Samsung's AI Innovation Program Member | Speaker |...

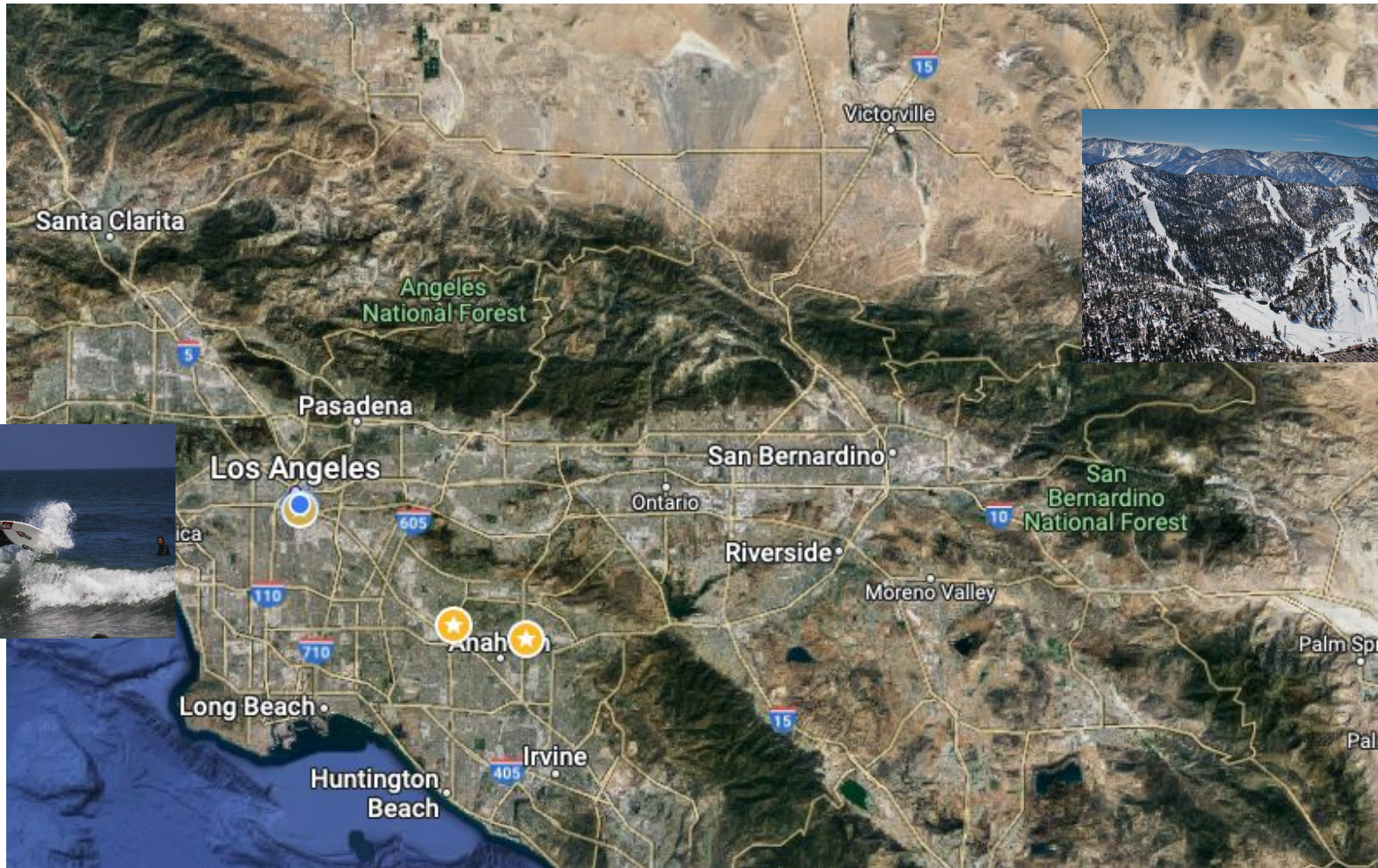
25 articles

+ Follow

September 2, 2023

📖 Open Immersive Reader

In today's fast-paced world, maintaining a balanced and nutritious diet can be a



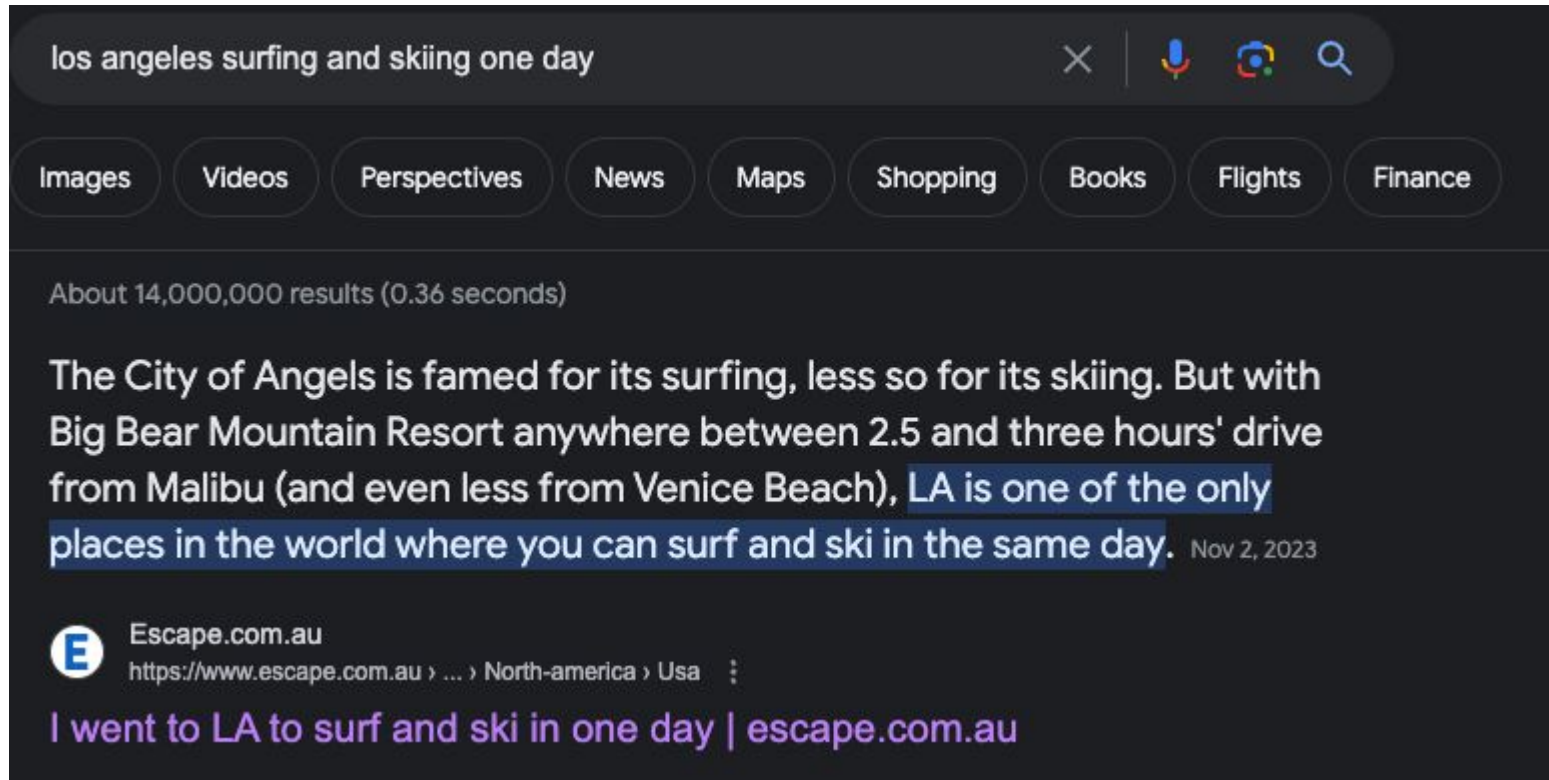
Can LLM automatically plan?

Prompt: “i am living in los angeles and today is january 11th. i want to do surf and ski in one day tomorrow. could you give me plan for me with time line?”

Can LLM automatically plan?

Prompt: “i am living in **malibu** and today is january 11th. i want to do surf and ski in one day tomorrow. could you give me plan for me with time line?”

Can LLM automatically plan?




los angeles surfing and skiing one day

Images Videos Perspectives News Maps Shopping Books Flights Finance

About 14,000,000 results (0.36 seconds)

The City of Angels is famed for its surfing, less so for its skiing. But with Big Bear Mountain Resort anywhere between 2.5 and three hours' drive from Malibu (and even less from Venice Beach), LA is one of the only places in the world where you can surf and ski in the same day. Nov 2, 2023

 Escape.com.au
<https://www.escape.com.au> › ... › North-america › Usa

I went to LA to surf and ski in one day | [escape.com.au](https://www.escape.com.au)

Can LLM automatically plan?

LLM is very good at ~~planning~~ retrieving
knowledge and approximating answers

Can LLM automatically plan?

“**planning** is an exploration to decide what **actions** need to be taken **to achieve a given goal.**”

HOW DO HUMANS DO PLANNING?

How do we do planning?



Goal

How do we do planning?

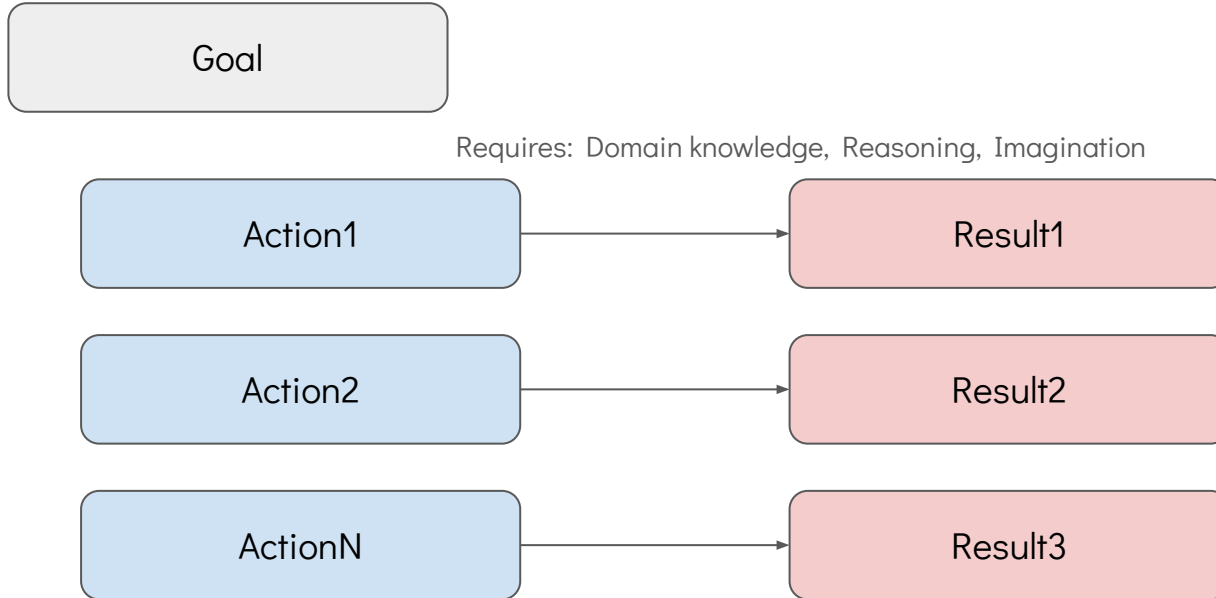
Goal

Action1

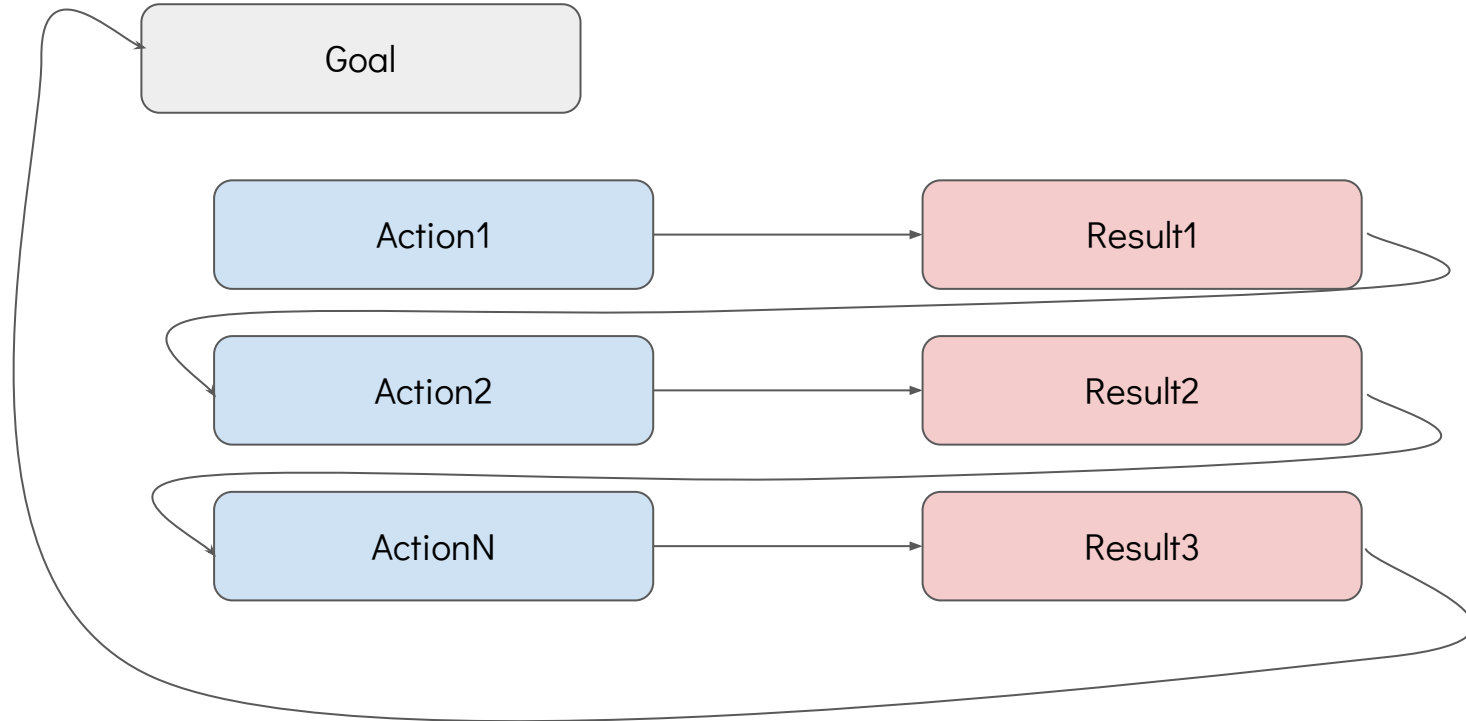
Action2

ActionN

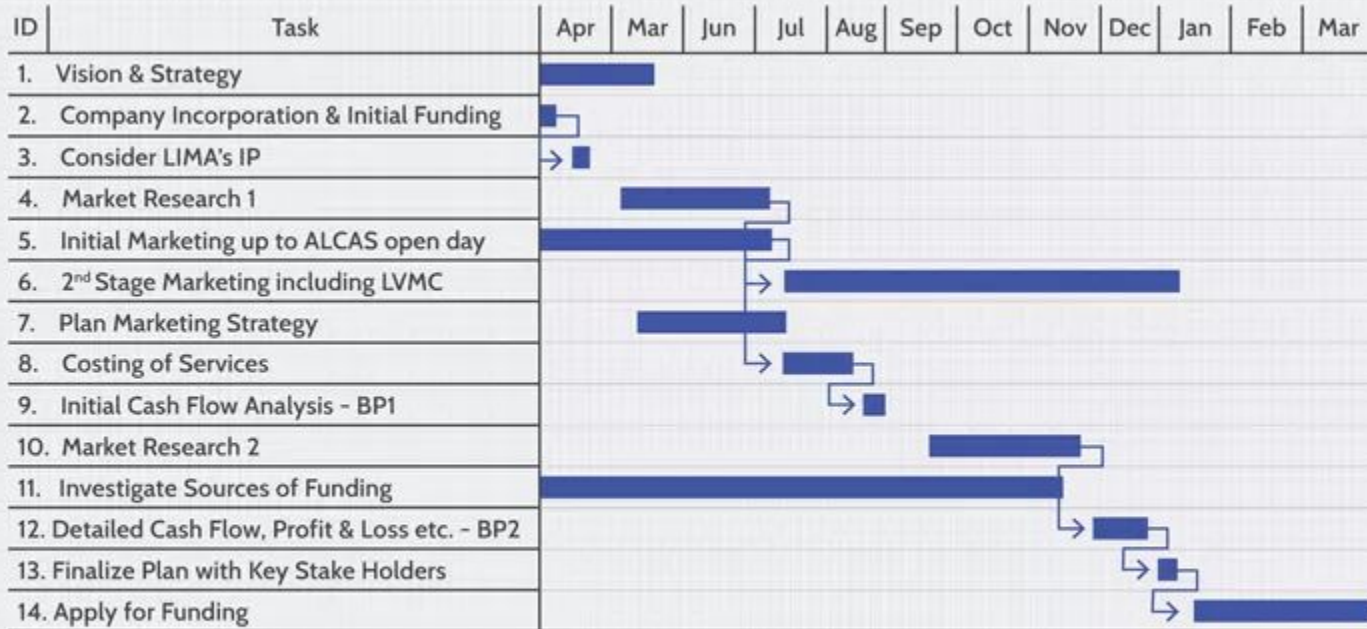
How do we do planning?



How do we do planning?



How do we do planning?



SEARCH AUGMENTED PLANNING

Game of 24

- You are given 4 numbers
- Find equation to reach 24

Game of 24

- [6, 6, 6, 6]
- ?

Game of 24

- [6, 6, 6, 6]
- $6 + 6 + 6 + 6 = 24$

Game of 24

- [4, 5, 6, 10]
-

Game of 24

- [4, 5, 6, 10]
- $5 * 6 + 4 - 10 = 24$

Game of 24

- Can LLM do this?

Game of 24

- Naive Prompt

Use numbers and basic arithmetic operations (+ - * /) to obtain 24.

Input: 4 5 6 10

Answer:

Game of 24

- Naive Prompt

Use numbers and basic arithmetic operations (+ - * /) to obtain 24.

Input: 4 5 6 10

Use numbers and basic arithmetic operations (+ - * /) to

Santamonica ▾

Send

4+5*6+10

Game of 24

- Few-Shot Prompt

Use numbers and basic arithmetic operations (+ - * /) to obtain 24.

Input: 4 4 6 8

Answer: $(4 + 8) * (6 - 4) = 24$

Input: 2 9 10 12

Answer: $2 * 12 * (10 - 9) = 24$

Input: 4 9 10 13

Answer: $(13 - 9) * (10 - 4) = 24$

Input: 1 4 8 8

Answer: $(8 / 4 + 1) * 8 = 24$

Input: 5 5 5 9

Answer: $5 + 5 + 5 + 9 = 24$

Input: 4 5 6 10

Answer:

Game of 24

- Few-Shot Prompt

Use numbers and basic arithmetic operations (+ - * /) to obtain 24.

Input: 4 4 6 8

Answer: $(4 + 8) * (6 - 4) = 24$

Input: 2 9 10 12

Use numbers and basic arithmetic operations (+ - * /) to

Santamonica ▾

Send

4 * 5 * 6 * 10 = 240, 240 / 10 = 24

Input: 5 5 5 9

Answer: $5 + 5 + 5 + 9 = 24$

Input: 4 5 6 10

Answer:

Game of 24

- Few-Shot + CoT Prompt

Use numbers and basic arithmetic operations (+ - * /) to obtain 24. Each step, you are only allowed to choose two of the remaining numbers to obtain a new number.

Input: 4 4 6 8

Steps:

$$4 + 8 = 12 \text{ (left: 4 6 12)}$$

$$6 - 4 = 2 \text{ (left: 2 12)}$$

$$2 * 12 = 24 \text{ (left: 24)}$$

$$\text{Answer: } (6 - 4) * (4 + 8) = 24$$

Input: 2 9 10 12

Steps:

$$12 * 2 = 24 \text{ (left: 9 10 24)}$$

$$10 - 9 = 1 \text{ (left: 1 24)}$$

$$24 * 1 = 24 \text{ (left: 24)}$$

$$\text{Answer: } (12 * 2) * (10 - 9) = 24$$

Input: 4 9 10 13

Steps:

$$13 - 10 = 3 \text{ (left: 3 4 9)}$$

$$9 - 3 = 6 \text{ (left: 4 6)}$$

$$4 * 6 = 24 \text{ (left: 24)}$$

$$\text{Answer: } 4 * (9 - (13 - 10)) = 24$$

Input: 1 4 8 8

Game of 24

- Few-Shot + CoT Prompt

Use numbers and basic arithmetic operations (+ - * /) to obtain 24. Each step, you are only allowed to choose two of the remaining numbers to obtain a new number.

Input: 4 4 6 8

Steps:

$4 + 8 = 12$ (left: 4 6 12)

$6 - 4 = 2$ (left: 2 12)

Use numbers and basic arithmetic operations (+ - * /) to

Santamonica ▾

Send

$(4 + 5) * (6 + 10) = 24$

$24 * 1 = 24$ (left: 24)

Answer: $(12 * 2) * (10 - 9) = 24$

Input: 4 9 10 13

Steps:

$13 - 10 = 3$ (left: 3 4 9)

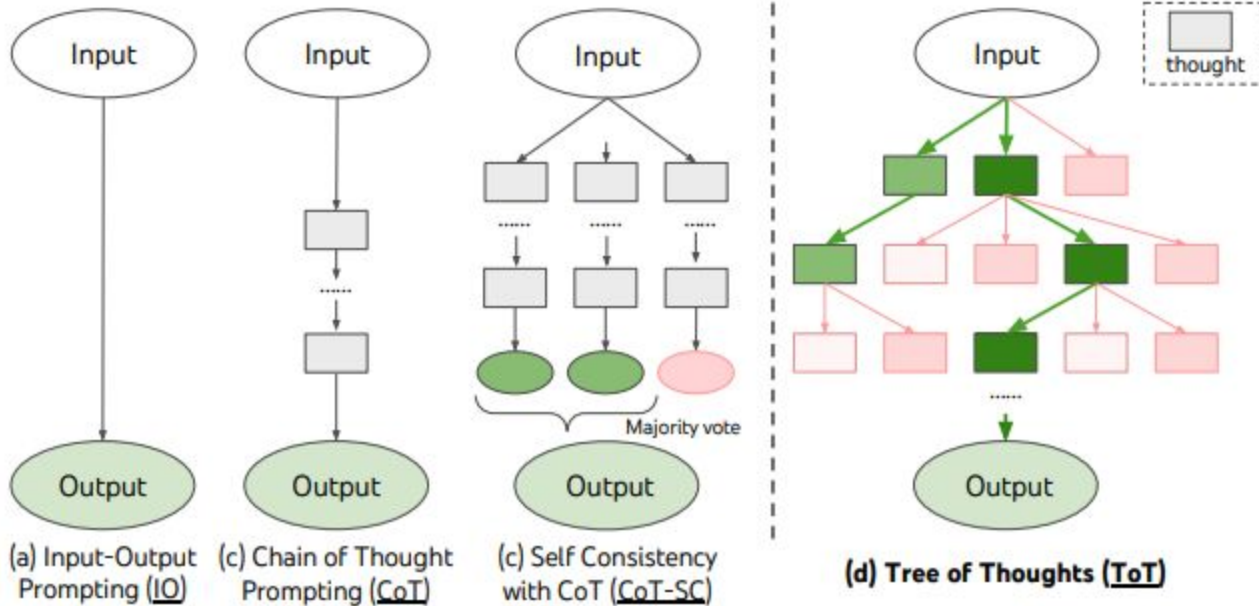
$9 - 3 = 6$ (left: 4 6)

$4 * 6 = 24$ (left: 24)

Answer: $4 * (9 - (13 - 10)) = 24$

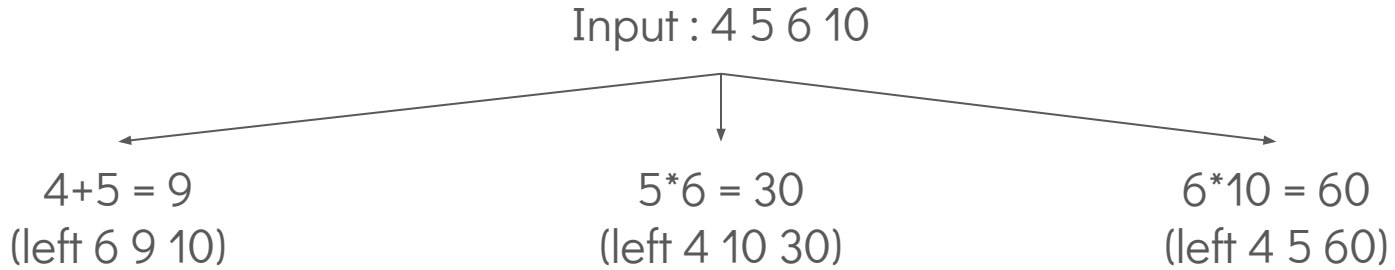
Input: 1 4 8 8

Tree-of-Thought



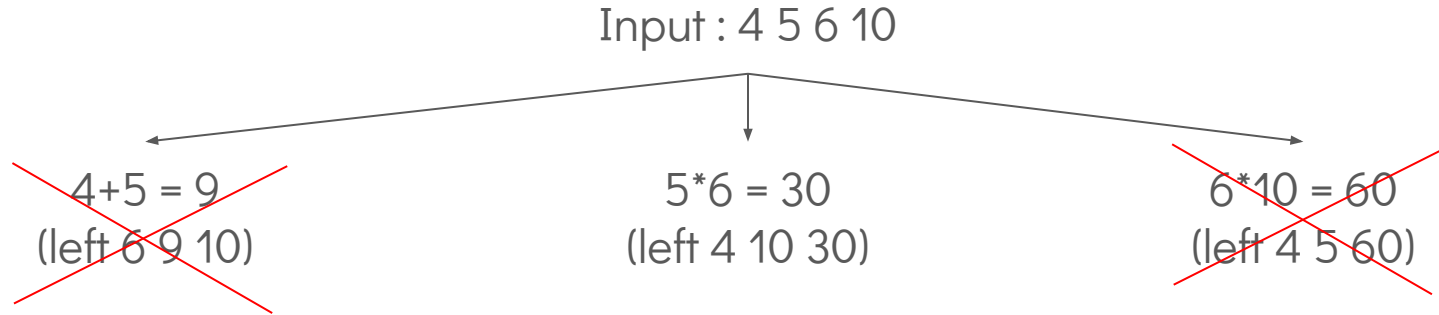
Tree-of-Thought

“Propose Candidates”



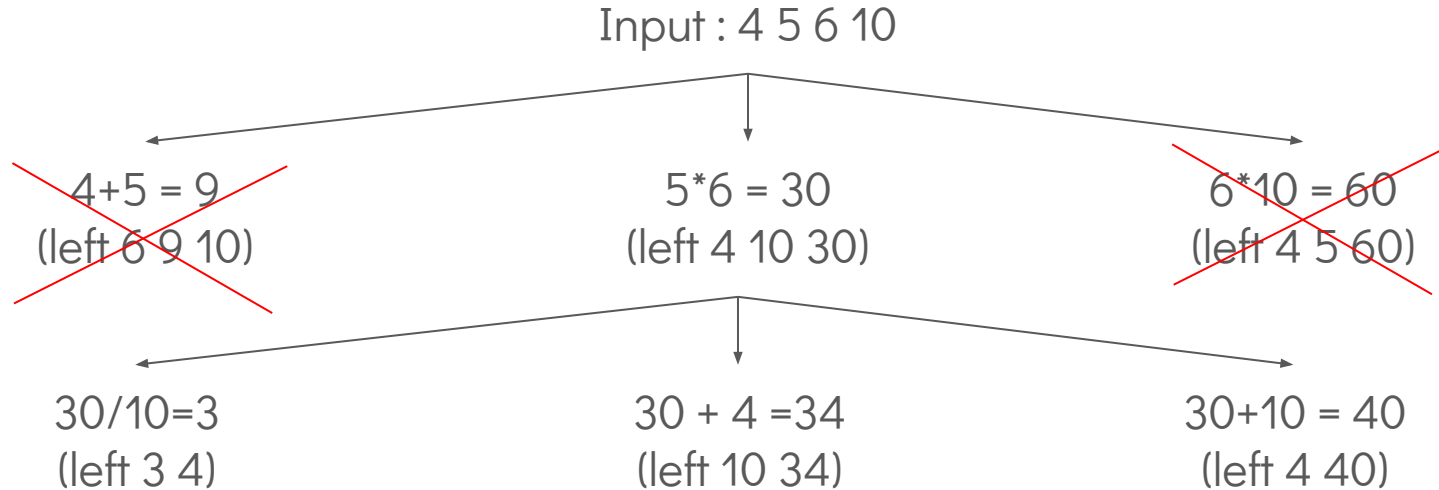
Tree-of-Thought

“Value Candidates”



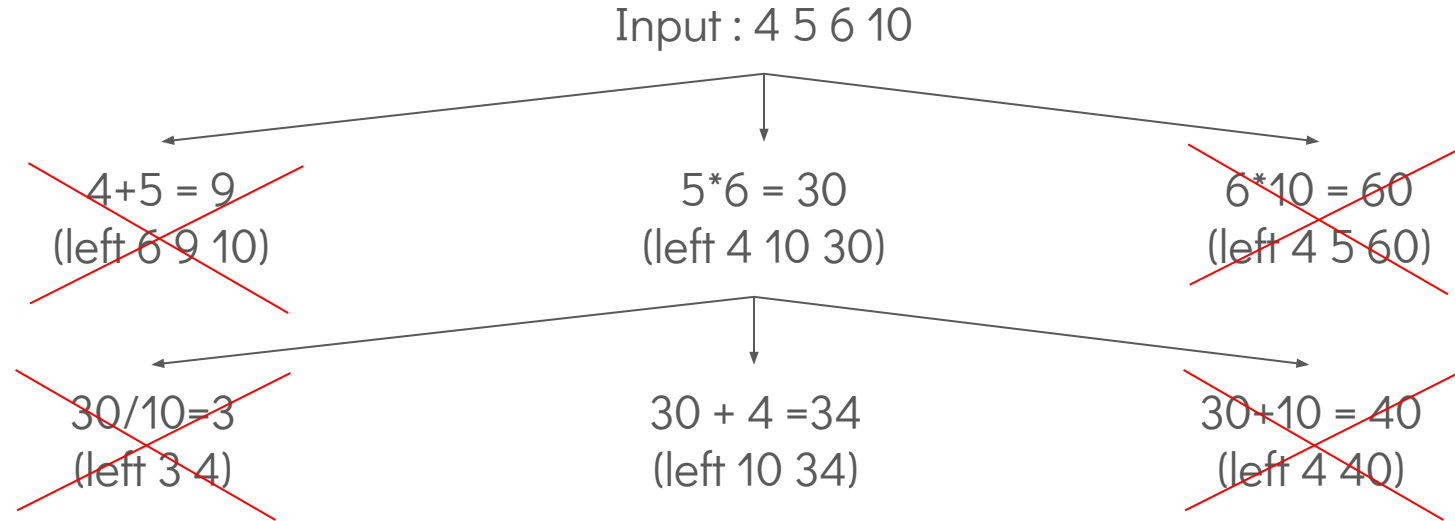
Tree-of-Thought

“Propose Candidates”

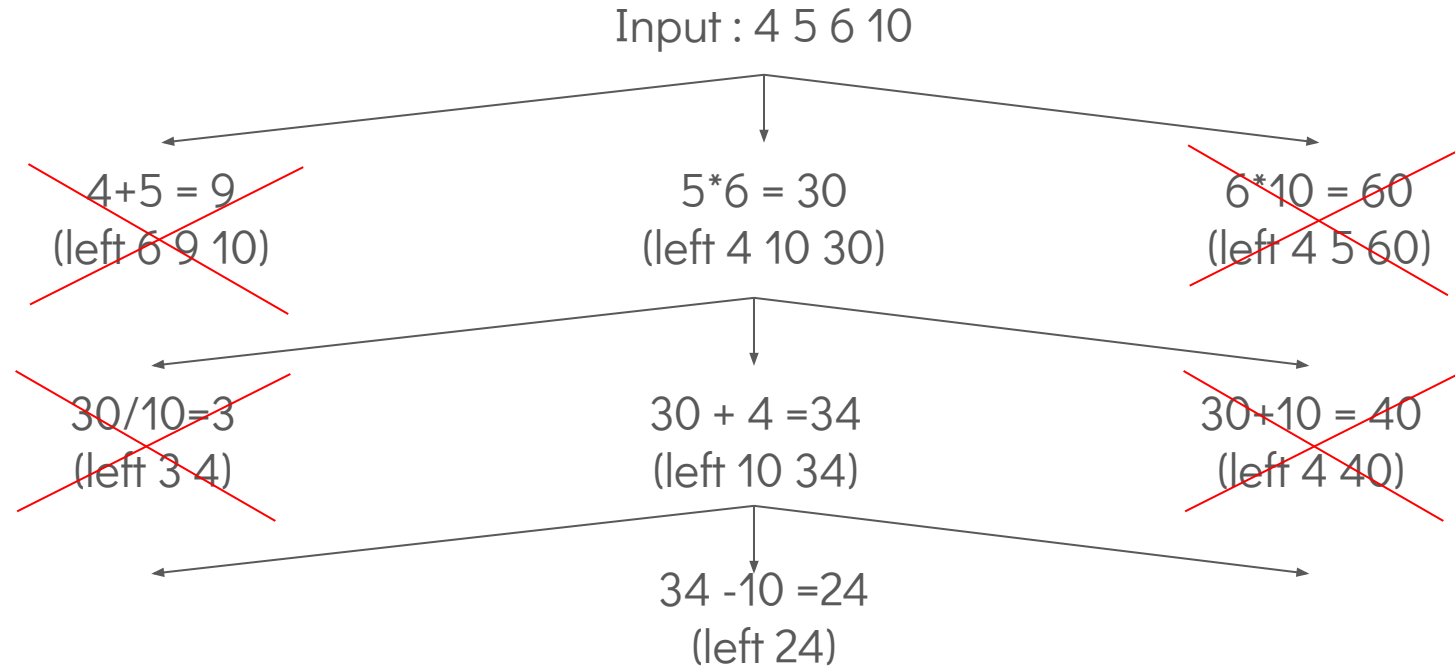


Tree-of-Thought

“Value Candidates”



Tree-of-Thought



Tree-of-Thought

```
-- new_ys --: ('4 + 5 = 9 (left: 6 9 10)\n', '5 + 6 = 11 (left: 4 10 11)\n', '6 - 4 = 2 (left: 2 5 10)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n', '10 - 4 = 6 (left: 5 6 6)\n', '10 + 4 = 14 (left: 5 6 14)\n', '4 * 5 = 20 (left: 6 10 20)\n', '10 - 5 = 5 (left: 4 5 6)\n')
-- sol values --: (3.0, 3.0, 3.0, 3.0, 3.0, 2.001, 2.001, 2.001)
-- choices --: ['4 + 5 = 9 (left: 6 9 10)\n', '5 + 6 = 11 (left: 4 10 11)\n', '6 - 4 = 2 (left: 2 5 10)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n', '10 - 4 = 6 (left: 5 6 6)\n']
```


Tree-of-Thought

```
-- new_ws --: ('10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 - 6 = 24 (left: 24)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 / 6 = 5 (left: 5)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n6 * 30 = 180 (left: 180)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 + 6 = 36 (left: 36)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n2.5 * 2 = 5 (left: 5 5 6)\n5 + 5 = 10 (left: 6 10)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n2.5 * 2 = 5 (left: 5 5 6)\n5 * 5 = 25 (left: 6 25)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n2.5 * 2 = 5 (left: 5 5 6)\n6 - 5 = 1 (left: 1 5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n2.5 * 2 = 5 (left: 5 5 6)\n6 / 5 = 1.2 (left: 1.2 5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 - 2.5 = 3.5 (left: 3.5 5 5)\n3.5 + 5 = 8.5 (left: 5 8.5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 - 2.5 = 3.5 (left: 3.5 5 5)\n5 - 1.5 (left: 1.5 5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 - 2.5 = 3.5 (left: 3.5 5 5)\n5 / 3.5 = 1.4 (left: 1.4 5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 / 2.5 = 2.4 (left: 2.4 5 5)\n2.4 + 5 = 7.4 (left: 5 7.4)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 / 2.5 = 2.4 (left: 2.4 5 5)\n5 + 2.4 = 7.4 (left: 5 7.4)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n5 - 2.5 = 2.5 (left: 2.5 2.5 6)\n2.5 + 2.5 = 5 (left: 5 6)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n5 - 2.5 = 2.5 (left: 2.5 2.5 6)\n2.5 * 2.5 = 6.25 (left: 6 6.25)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n5 - 2.5 = 2.5 (left: 2.5 2.5 6)\n6 - 2.5 = 3.5 (left: 2.5 3.5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n5 - 2.5 = 2.5 (left: 2.5 2.5 6)\n6 / 2.5 = 2.4 (left: 2.4 2.5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 - 2.5 = 3.5 (left: 3.5 5 5)\n3.5 * 5 = 17.5 (left: 5 17.5)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n6 / 2.5 = 2.4 (left: 2.4 5 5)\n5 - 2.4 = 2.6 (left: 2.6 5)\n')
-- sol values --: (60.0, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.002, 0)
-- choices --: ['10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 - 6 = 24 (left: 24)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 / 6 = 5 (left: 5)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n6 * 30 = 180 (left: 180)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 + 6 = 36 (left: 36)\n', '10 / 4 = 2.5 (left: 2.5 5 6)\n2.5 * 2 = 5 (left: 5 5 6)\n5 + 5 = 10 (left: 6 10)\n']
```



Tree-of-Thought

```
['10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 - 6 = 24 (left: 24)\nAnswer: (5 * (10 - 4)) - 6 = 24\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 / 6 = 5 (left: 5)\n5 + 2 = 7 (left: 7 8 8 14)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 / 6 = 5 (left: 5)\n5 - 2 = 3 (left: 3 8 8 14)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 / 6 = 5 (left: 5)\n14 / 5 = 2.8 (left: 2.8 2 8 8)\n', '10 - 4 = 6 (left: 5 6 6)\n5 * 6 = 30 (left: 6 30)\n30 / 6 = 5 (left: 5)\n14 - 5 = 9 (left: 9 2 8 8)\n']>>> print(ys[0])
10 - 4 = 6 (left: 5 6 6)
5 * 6 = 30 (left: 6 30)
30 - 6 = 24 (left: 24)
Answer: (5 * (10 - 4)) - 6 = 24
```

LLM AS LEARNER

Voyager(Minecraft)


Voyager: An Open-Ended Embodied Agent with Large Language Models

Guanzhi Wang ^{1 2}, Yuqi Xie³, [Yunfan Jiang](#)^{4*}, Ajay Mandlekar^{1*},
Chaowei Xiao^{1 5}, Yuke Zhu^{1 3}, Linxi "Jim" Fan ^{1†}, Anima Anandkumar^{1 2†}

¹NVIDIA, ²Caltech, ³UT Austin, ⁴Stanford, ⁵ASU

*Equal contribution †Equal advising

 Corresponding authors: guanzhi@caltech.edu, dr.jimfan.ai@gmail.com

 arXiv

 PDF

 Code

 Tweet

 MineDojo

Voyager(Minecraft)



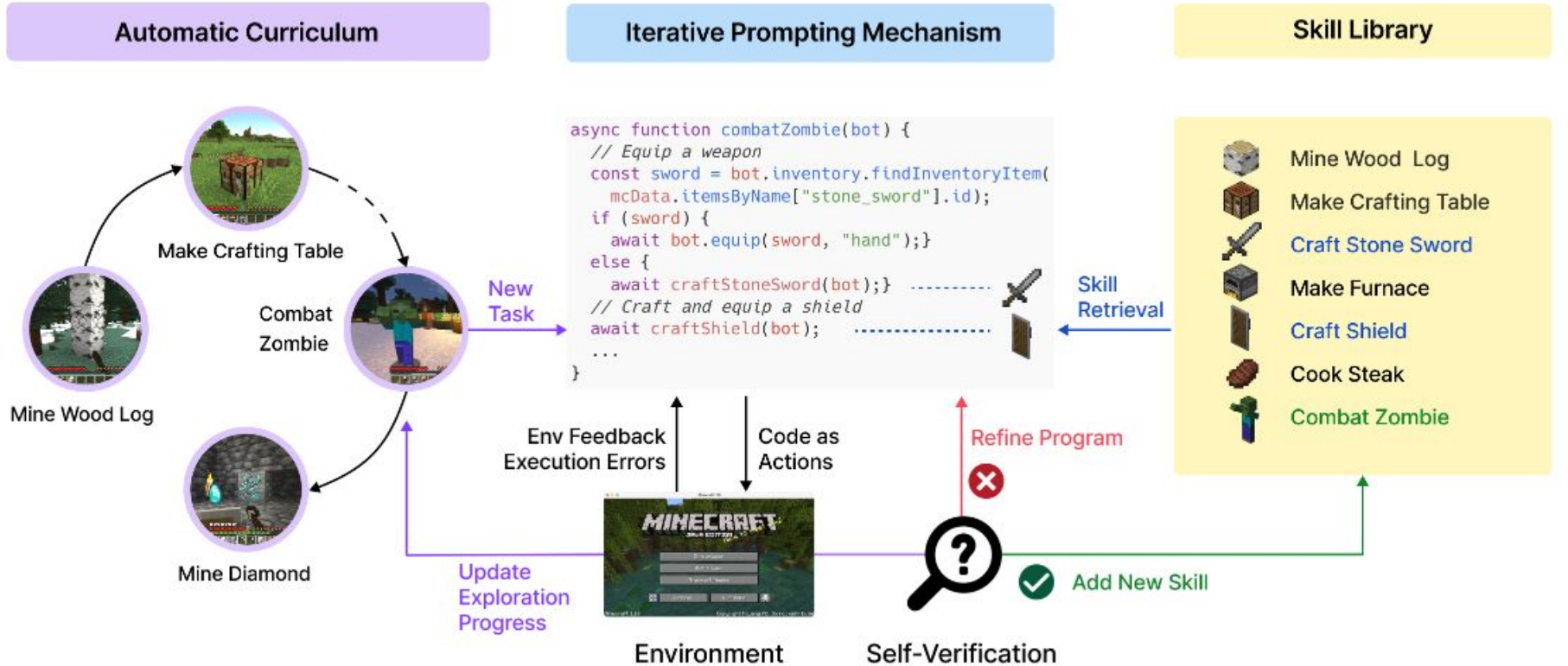
Voyager(Minecraft)



Voyager(Minecraft)



Voyager(Minecraft)



Voyager(Minecraft)

Automatic Curriculum



Inventory (5/36): {'oak_planks': 3, 'stick': 4, 'crafting_table': 1, 'stone': 3, 'wooden_pickaxe': 1}

GPT-4

Reasoning: Since you have a wooden pickaxe and some stones, it would be beneficial to upgrade your pickaxe to a stone pickaxe for better efficiency.
Task: Craft 1 stone pickaxe.



Biome: river

Inventory (4/36): {'sugar_cane': 3, 'paper': 3, 'fishing_rod': 1, 'wooden_pickaxe': 1}

GPT-4

Reasoning: Since you have a fishing rod in your inventory and you are near a river biome, it would be a good opportunity to catch some fish for food and experience.
Task: Catch 1 fish.



Nearby entities: pig, cat, villager

Health: 12/20

Hunger: 0/20

GPT-4

Reasoning: Your hunger is at 0, which means you need to find food to replenish your hunger. Since there are pigs nearby, you can kill one to obtain raw porkchops.
Task: Kill 1 pig.



Inventory (6/36): {'furnace': 1, 'stone_pickaxe': 1, 'oak_planks': 7, 'crafting_table': 1, 'raw_iron': 4, 'coal': 1}

GPT-4

Reasoning: You have raw iron and coal, and you have a furnace. It's time to smelt the iron to make iron ingots, which can be used to craft better tools and armor.
Task: Smelt 4 raw iron.



Time: night

Nearby entities: zombie

Equipment: 'stone_sword', 'shield'

GPT-4

Reasoning: Since it's night and there's a zombie nearby, it's a good opportunity to try killing the zombie now that you have a stone sword and a shield equipped.
Task: Kill 1 zombie.

Voyager(Minecraft)

Skill Library

Program Generated by GPT-4

```
async function combatZombie(bot) {  
  // Equip a weapon  
  const sword =  
  bot.inventory.findInventoryItem(  
    mcData.itemsByName[  
      "stone_sword"  
    ],id  
  );  
  if (sword) {  
    await bot.equip(sword, "hand");  
  } else {  
    await craftStoneSword(bot);  
    ...  
  }  
  // Craft and equip a shield  
  ...  
  // Recover hunger  
  ...  
  // Look for and combat a zombie  
  ...  
}
```

GPT-3.5

Program Description

```
async function combatZombie(bot) {  
  // The function is about  
  equpping a stone sword to combat  
  a zombie. If a stone sword is not  
  found, it will craft one.  
  Additionally, it crafts and equips  
  a shield for added protection.  
  Afterwards, it proceeds to cook  
  sticks in order to restore hunger.  
  Once hunger is replenished, it  
  actively searches for a zombie and  
  engages in combat with it.  
}
```

Embedding

Key

Value

Add

Skill Library

- Mine Wood Log
- Make Crafting Table
- Craft Wooden Pickaxe
- Craft Stone Sword
- Make Furnace
- ...
- Combat Cow
- Cook Steak
- Craft Iron Axe
- Combat Zombie

Task: Craft Iron Pickaxe

How to craft an iron pickaxe in
Minecraft?

GPT-3.5

To craft an iron pickaxe, you
need to 3 iron ingots and 2
sticks. Once you have gathered
the materials, ...

Environment Feedback

Embedding

Query

Skill Library

Retrieve

- Smelt Iron Ingot
- Craft Stick
- Make Crafting Table
- Make Furnace
- Craft Wooden Pickaxe

Voyager(Minecraft)

Iterative Prompting Mechanism

Environment Feedback

I cannot make stick because I need: 2 more planks
I cannot make stone_shovel because I need: 2 more stick

GPT-4

```
async function craftStoneShovelWithTable(bot) {  
  // If not enough cobblestone, mine cobblestone  
  ...  
  + // If not enough sticks, check if there are  
  + // enough planks in the inventory  
  if (sticksCount < 2) {  
  +   const planksCount = bot.inventory.count(  
  +     mcData.itemsByName.acacia_planks.id);  
  +   if (planksCount < 2) {  
  +     // Collect or craft planks  
  +     await mineBlock(bot, "acacia_log", 1);  
  +     await craftItem(bot, "acacia_planks", 1);  
  +   }  
  +   // Craft sticks using planks  
  +   await craftItem(bot, "stick", 1);  
  await craftItem(bot, "stone_shovel", 1);  
}
```

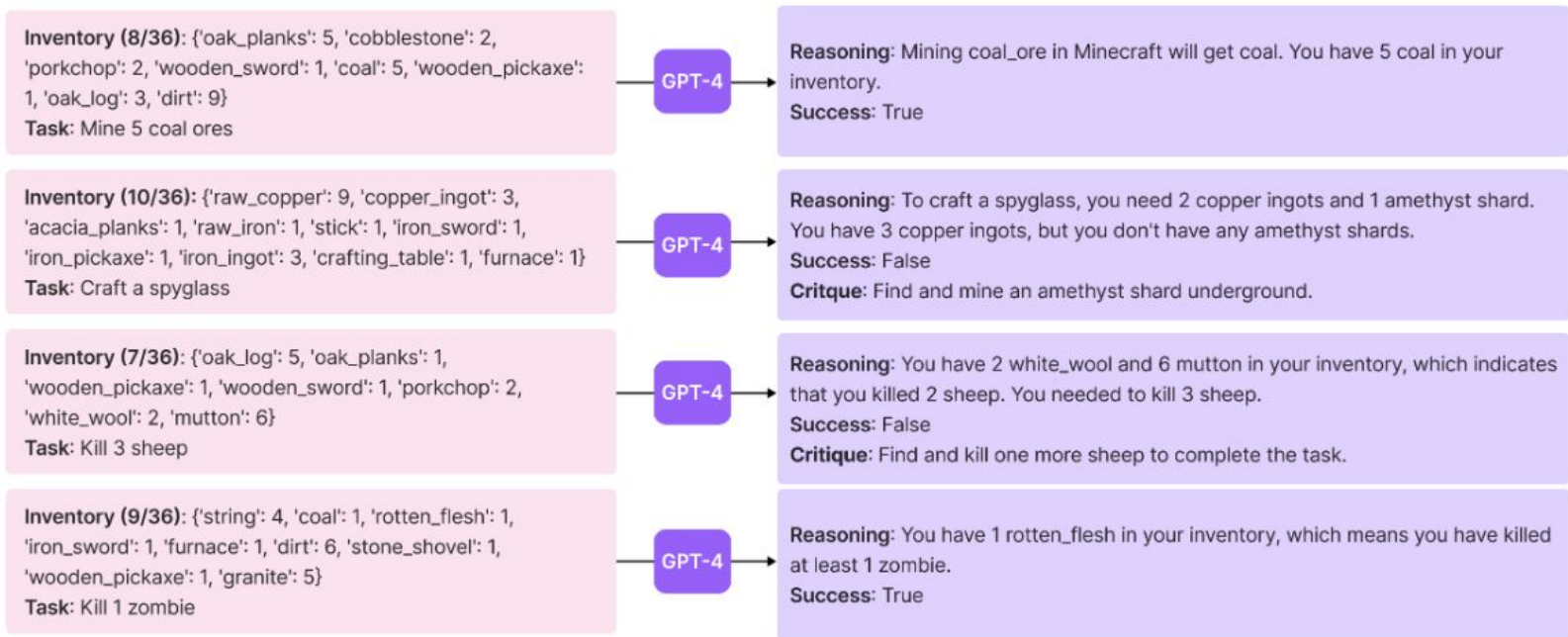
Execution Error

throw new Error(`No item named \${name}`);
No item named acacia_axe
at line 18:await craftItem(bot, "acacia_axe", 1);

GPT-4

```
-async function craftAcaciaAxe(bot) {  
+async function craftWoodenAxe(bot) {  
  // Craft anacia planks and sticks  
  ...  
  // Place the crafting table near the bot  
  ...  
- // Craft an acacia axe using 3 acacia planks  
- // and 2 sticks  
- await craftItem(bot, "acacia_axe", 1);  
- bot.chat("Acacia axe crafted.");  
+ // Craft a wooden axe using 3 acacia planks  
+ // and 2 sticks  
+ await craftItem(bot, "wooden_axe", 1);  
+ bot.chat("Wooden axe crafted.");  
}
```

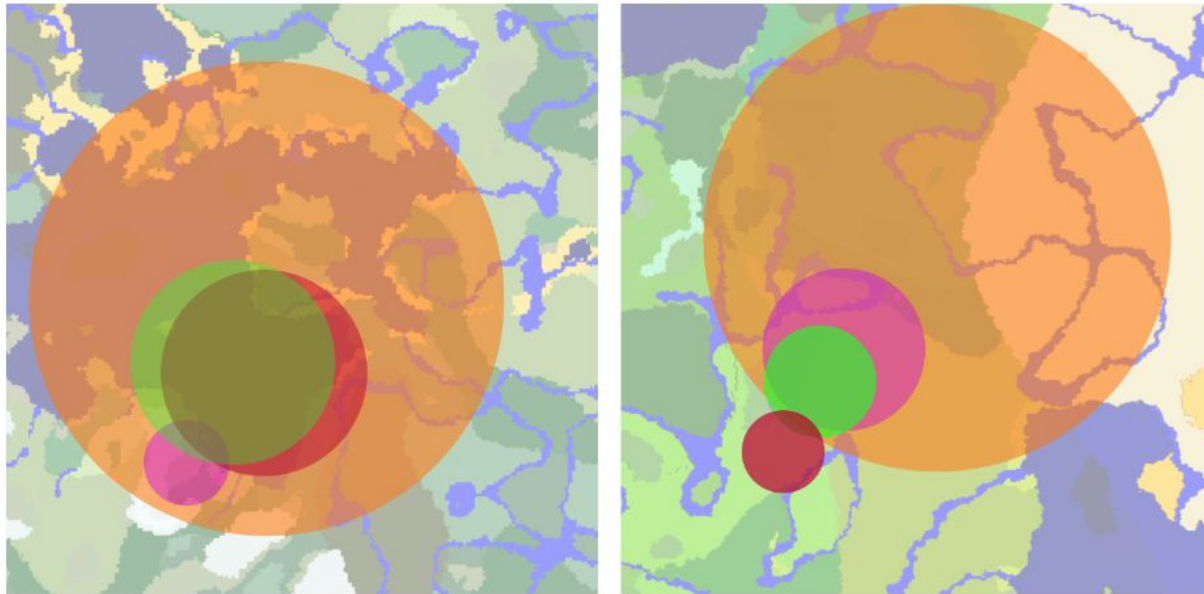
Voyager(Minecraft)



Self-verification. By providing the agent's current state and the task to GPT-4, we ask it to act as a critic and inform us whether the program achieves the task. In addition, if the task fails, it provides a critique by suggesting how to complete the task.

Voyager(Minecraft)

Extensive Map Traversal



—●— Voyager (Ours)

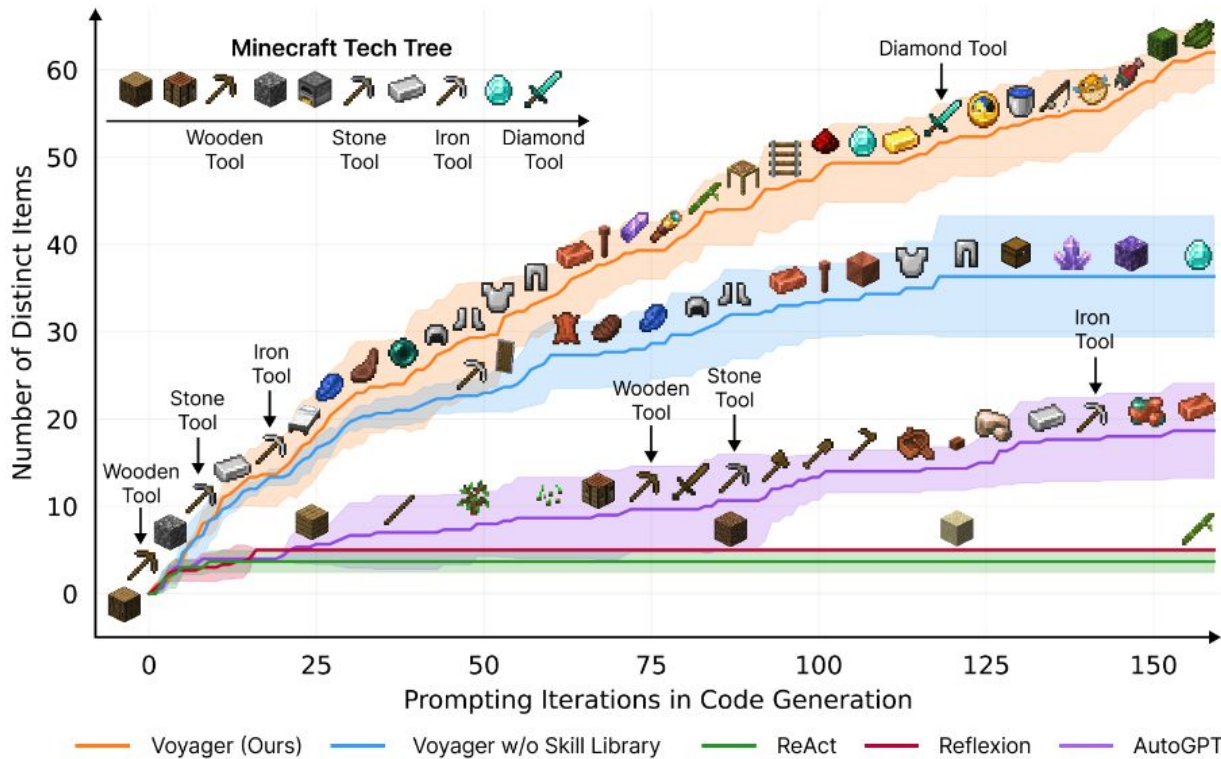
—●— ReAct

—●— Reflexion

—●— AutoGPT

Map coverage: Two bird's eye views of Minecraft maps. Voyager is able to navigate distances 2.3x longer compared to baselines by traversing a variety of terrains, while the baseline agents often find themselves confined to local areas, which significantly hampers their capacity to discover new knowledge.

Voyager(Minecraft)



Voyager discovers new Minecraft items and skills continually by self-driven exploration, significantly outperforming the baselines.

LLM AS TOOL RETRIEVER

LLM AS TOOL RETRIEVER

Toolformer: Language Models Can Teach Themselves to Use Tools

Timo Schick Jane Dwivedi-Yu Roberto Dessì[†] Roberta Raileanu
Maria Lomeli Luke Zettlemoyer Nicola Cancedda Thomas Scialom

Meta AI Research [†]Universitat Pompeu Fabra

The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

The name derives from "la tortuga", the Spanish word for [MT("tortuga") → turtle] turtle.

The Brown Act is California's law [WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.] that requires legislative bodies, like city councils, to hold their meetings open to the public.

LLM AS TOOL RETRIEVER

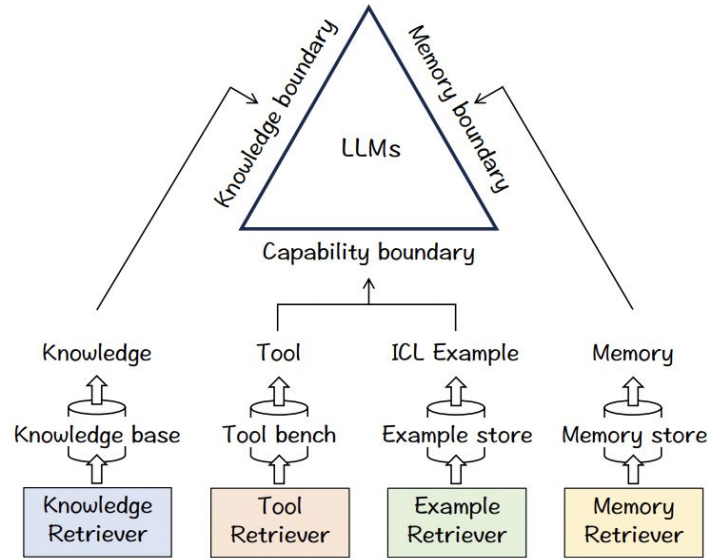


Figure 1: Confront the threefold inherent boundaries of LLMs on top of retrieval augmentation.

Zhang, Peitian, et al. "Retrieve anything to augment large language models." arXiv preprint arXiv:2310.07554 (2023).

LLM AS TOOL RETRIEVER



LLM AS TOOL RETRIEVER



Secret Santa
Questionnaire

 Name: _____

Some of my favorite things:

Snack: _____	Coffee or Tea
Color: _____	Candy or Chocolate
Restaurant: _____	Sweet or Salty
Scent: _____	Vanilla or Chocolate
Hobby: _____	Go out or Stay in
Store: _____	Alcohol or Non-Alcoholic
Drink: _____	Brights or Neutrals
TV Show: _____	Books or Movies
Sports Team: _____	Shopping online or in store
Candy: _____	Cats or Dogs

Other things I LOVE: _____

Things I don't like: _____

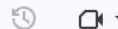
LLM AS TOOL RETRIEVER



LLM AS TOOL RETRIEVER

2022 (Responses) ☆ 🔄 ☁

t Format Data Tools Extensions Help



10% ▾ View only

D	E	F	G	H	I	J	K	L	M
Favorite foods	Drink of Choice?	What types of books do you like?	Websites, subreddits, youtube channels that speak to me	Favorite snack/candy/indulgence	Coffee or Tea?	11AM on a Sunday, I am probably _____.	What hobbies/activities are you into or would like to get in to?	Some of my favorite local spots:	I like to buy _____ for home:
dried mango	unique waters	biographies			I enjoy both	at a playground	pilates	anything new/weird in the arts district	not sure
Beef jerky	Coffee	Science & engineering, travel & adventure non-fiction, weather, space, math	Adam Ragusea, NileRed, Applied Science, /r/clouds, /r/socalthiking, /r/chess, /r/hamradio, /r/sailing	Beef jerky	Coffee enthusiast	At a coffee shop or on a hike	Hiking, woodworking, chess, sailing, comedy. Getting into dnd thanks to Kevin.	Yang's Kitchen, Eightfold Coffee, Terra Mia Coffee	Toilet paper
pizza	craft ipa	Sci fi			Coffee enthusiast	playing guitar	writing music	Native Son	
ramen, chocolate	aperol spritz, chartreuse cocktails	any classic novel i read very few of them :(ML twitter, yannic kilcher, those yt channels with cool visuals for mathy things, standup channels, gopro videos with ski jumps i'd break a leg trying	After Eight, pie hole	Coffee enthusiast	sleeping	na	afuri, 2 bit circus, broad, la brea pits, santa monica	food, furniture, cc tea
Zamn, uh, sweet, but not too sweet, or spicy. No mush (oatmeal, grits, etc.).	Tonic, or uhhhh grape Fanta?	Stuff about constructing languages, or about interesting linguistic theories or about cool coding stuff (like the garbage collection <- ref counting boek)	You don't want to know. (Music stuff, plugin reviews (sometimes), um vtubers?) https://www.youtube.com/watch?v=S4nwKl_xF	Caramel, or um sour gobsmackers, or like something sweet and sour	Tea enthusiast	lounging in bed or programming another project I will never finish	Music Production TM, id want to draw, but that needs time and practice, and I have neither	You're supposed to have these? umm like umm, my apartment. srsly, erm maybe Salt & Straw?	furniture, electronic music gadgets
	ou erh tea (butter			turtle chips sweet corn			visualizing 4-manifolds		

LLM AS TOOL RETRIEVER



Carin Rogoff <carin@gridspace.com>

to bcc: me ▾

You should have received your **secret santa** assignments now!

Aim for roughly \$20-\$30. We will open presents at the office on Dec 15th, the day of the holiday party. Most people have asked to have their gift sent to the office, but make sure to check in the responses. We encourage you to wrap your presents!

Learn about your **secret santa** assignee [here](#):
Fill out the [survey](#) if you haven't yet.

LLM AS TOOL RETRIEVER

Please set up secret santa for everyone. you can find their phone number from the directory. ask them the following questions: 1. Drink of Choice? 2. What types of books do you like? 3. Websites, subreddits, youtube channels that speak to me, 4. Favorite snack/candy/indulgence 5. Coffee or Tea?. 6. What hobbies/activities are you into or would like to get in to? 7. Games I like: 8. Tshirt size 9. Please don't get me _____. 10. I can never have too many _____. Make calls in parallel. For each participant, prompt once and wait for their response before any follow up questions. record their responses in a file and keep it updated. If the response is incomplete, make a few attempts to follow up.

Can you send a reminder to Hagid and Tom to ask for their input? let them know we will assign secret santa by the end of today.

Can you remind Tom again?

review their preferences, assign each of them another person to be the gift sender. then contact the sender about who their recipient is, and what do they prefer for their gift. do not reveal to the recipient who their sender is. Tell them: "Aim for roughly \$20-\$30. We will open presents at the office on Dec 14th, the day of the holiday party. We encourage you to wrap your presents!"

Based on the preferences provided, I will now assign each participant another person to be their gift sender. After assigning, I will contact each sender about their recipient and their preferences.

To bottom

Type here.

Send



Return: *Hermione: I've never read them, but would be interested!!*

```
say(phone_number=+18644206681, message=Hermione me...
```

```
wait()
```

Return: *Albus: Dude kaladin is so sick in the way of kings, do you think I can become a lighteyes*

```
say(phone_number=+18644206681, message=Kaladin is ...
```

```
wait()
```

Return: *Harry: What's a good gift to give Draco?*

```
say(Considering Draco's preferences for IPA Beer, ...
```

Return: *Error: say() takes 2 positional arguments but 679 were given*

```
say({"phone_number":"+14243329752", "message":"Cons...
```

```
wait()
```

Return: *Harry: Who was picked to be my secret Santa?*

```
say(I'm sorry, Harry, but the identity of your Sec...
```

Return: *Error: say() takes 2 positional arguments but 297 were given*

```
say(phone_number=+14243329752, message=I'm sorry, ...
```

```
wait()
```

Return: *Harry: What would happen if I texted you a picture?*

Exercises (From last lecture)

- Hallucinations are a major problem with LLM generations. What are strategies for maximizing truthfulness of model outputs?

There are many approaches, but some include multiple LLM passes, prompting the generation to only reference source documents, forcing the model to cite input sources, and, even, telling the model to not hallucinate. From the planning perspective, recursive prompting can significantly reduce hallucinations.

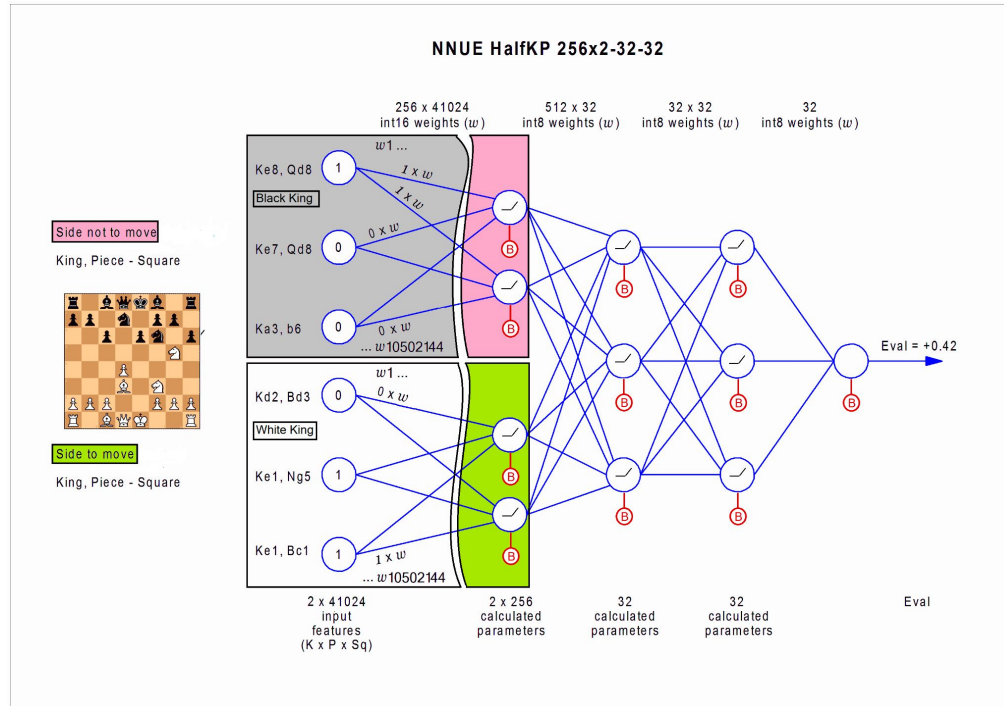
- How might you align an LLM to play a novel abstract strategy game using only a description of the rules?

While this is, in the general case, an unsolved problem. LLM's are going to perform much better on general game playing if an external system splits the problem into substeps such as turning the rules into a state space and the use of tree search.

Exercises (From last lecture)

- Stockfish is currently the most powerful chess engine by rating. Based on your own research, how much of this power can be attributed to machine learning versus non-ML algorithms.

The DNN NNUE contributes around 100 to the ELO of Stockfish.



Exercises

- In real-world, life does not always go planned. How LLM can address this problem when LLM does planning?
- Some approaches of LLM planning, list many actions and evaluate them and select. It might require huge computation, are there any way to reduce this waste of computation?

REFERENCES

- Mattar, Marcelo G, and Máté Lengyel. "Planning in the brain." *Neuron* vol. 110,6 (2022): 914-934
- Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in Neural Information Processing Systems* 35 (2022): 24824-24837.
- Yao, Shunyu, et al. "Tree of thoughts: Deliberate problem solving with large language models." *arXiv preprint arXiv:2305.10601* (2023).
- Schick, Timo, et al. "Toolformer: Language models can teach themselves to use tools." *arXiv preprint arXiv:2302.04761* (2023).
- Wang, Guanzhi, et al. "Voyager: An open-ended embodied agent with large language models." *arXiv preprint arXiv:2305.16291* (2023).